Hidden Markov Model

CS 470 Introduction To Artificial Intelligence

Daqing Yi

Department of Computer Science
Brigham Young University

Outline

Outline

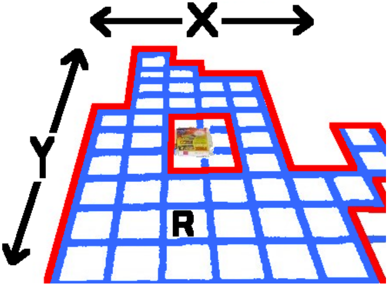
Hidden Markov Model

- hidden states
- observations
- Markov assumption

Application

Robot movement

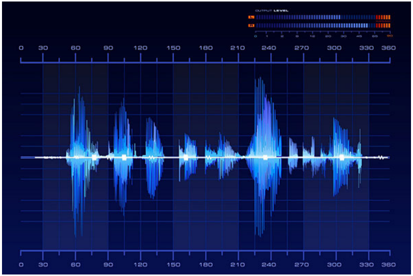
- hidden states : range sensor, visual sensor
- observations : location



Application

Speech processing

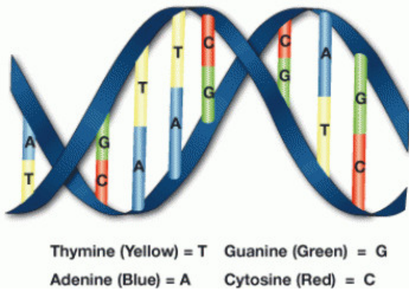
- hidden states : sound signal
- observations : parts of speech, words



Application

DNA sequence analysis

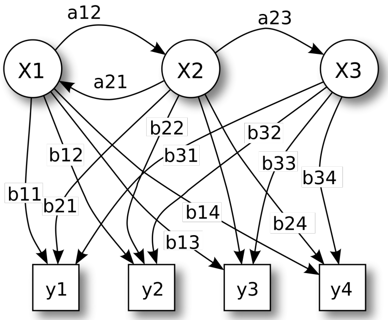
- hidden states : DNA base pairs
- observations : Genes



Outline

Hidden Markov Model

- a set of states $\{1, \cdots, S\}$
- transition model $P(X_t \mid X_{t-1})$
  - first-order Markov assumption $P(X_t \mid X_{0:t-1}) = P(X_t \mid X_{t-1})$
- a set of evidence Σ
- observation model $P(E_t \mid E_t)$
  - first-order Markov assumption $P(E_t \mid X_{0:t}, E_{0:t-1}) = P(E_t \mid X_t)$



Linear form

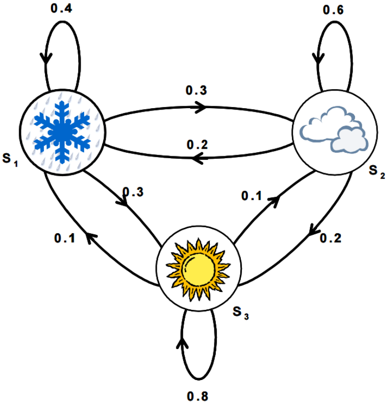
Example

- State vector

$$\boldsymbol{x}_t = \begin{bmatrix} \text{precipitation} \\ \text{cloudy} \\ \text{sunny} \end{bmatrix}$$

- Transition matrix

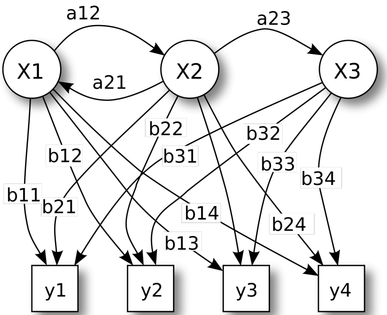
$$\boldsymbol{T} = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$



Application

Question?

- What is the value of current state?
- What is the probability that the next state is $X1$?
- What is the probability that the observation of the next state is $y3$?
- What is the most likely value of the next state?



HMM problems

- Probability evaluation
  - $P(E \mid X)$ or $P(X \mid E)$
  - Forward and Backward Algorithms
- Optimal state sequence
  - $\arg\max_X P(X \mid E)$
  - Viterbi Algorithm
- Parameter estimation
  - $\arg\max_\Lambda P(\lambda \mid E)$
  - Baum-Welch Algorithm

Outline

Forward Algorithm

filtering and prediction

$$
\begin{aligned}
\boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) =& \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t}, \boldsymbol{e}_{t+1}) \\
=& \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}, \boldsymbol{e}_{1:t}) \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t}) \\
=& \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t}) \\
=& \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \sum_{\boldsymbol{x}_t} \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{x}_t) P(\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})
\end{aligned}
$$

Forward Algorithm

- Let $\boldsymbol{f}_{1:t}$ be $\boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$
- "Message" propogation
- $\boldsymbol{f}_{1:t+1} = \alpha \textsc{Forward}(\boldsymbol{f}_{1:t}, \boldsymbol{e}_{t+1})$

Forward Algorithm

Algorithm

Priori $P(\boldsymbol{x}_1)$

- Initialization $\boldsymbol{P}(\boldsymbol{x}_1 \mid \boldsymbol{e}_1) = \alpha \boldsymbol{P}(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) P(\boldsymbol{x}_1)$
- Induction $\boldsymbol{P}(\boldsymbol{x}_{t+1} \mid \boldsymbol{e}_{1:t+1}) = \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{x}_{t+1}) \sum_{\boldsymbol{x}_t} P(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t) P(\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})$

Linear form

- $\boldsymbol{f}_{1:t} = \boldsymbol{P}(\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$
- $\boldsymbol{f}_{1:t+1} = \alpha \boldsymbol{O}_{t+1} \boldsymbol{T}^T \boldsymbol{f}_{1:t}$

Backward Algorithm

smoothing

$$
\begin{aligned}
\mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:t}) =& \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k}, \mathbf{e}_{k+1:t}) \\
=& \alpha \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k}) \mathbf{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k, \mathbf{e}_{1:k}) \\
=& \alpha \mathbf{P}(\mathbf{X}_k \mid \mathbf{e}_{1:k}) \mathbf{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k)
\end{aligned}
$$

Backward Algorithm

$$
\begin{aligned}
\boldsymbol{P}(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k) &= \sum_{\mathbf{x}_{k+1}} \boldsymbol{P}(\mathbf{e}_{k+1:t}, \mathbf{x}_{k+1} \mid \mathbf{X}_k) \\
&= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1:t} \mid \mathbf{X}_k, \mathbf{x}_{k+1}) \boldsymbol{P}(\mathbf{x}_{k+1} \mid \mathbf{X}_k) \\
&= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1:t} \mid \mathbf{x}_{k+1}) \boldsymbol{P}(\mathbf{x}_{k+1} \mid \mathbf{X}_k) \\
&= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1}, \mathbf{e}_{k+2:t} \mid \mathbf{x}_{k+1}) \boldsymbol{P}(\mathbf{x}_{k+1} \mid \mathbf{X}_k) \\
&= \sum_{\mathbf{x}_{k+1}} P(\mathbf{e}_{k+1} \mid \mathbf{x}_{k+1}) P(\mathbf{e}_{k+2:t} \mid \mathbf{x}_{k+1}) \boldsymbol{P}(\mathbf{x}_{k+1} \mid \mathbf{X}_k)
\end{aligned}
$$

Backward Algorithm

- Let $\boldsymbol{b}_{k+1:t}$ be $\boldsymbol{P}(\boldsymbol{b}_{k+2:t}, \boldsymbol{e}_{k+1})$
- "Message" backpropogation
- $\boldsymbol{b}_{k+1:t} = \alpha\textsc{Backward}(\boldsymbol{b}_{k+2:t}, \boldsymbol{e}_{k+1})$

Backward Algorithm

Algorithm

- Initialization $\boldsymbol{P}(\Phi \mid \boldsymbol{x}_t) = 1$
- Induction $\boldsymbol{P}(\boldsymbol{e}_{k+1:t} \mid \boldsymbol{x}_k) = \sum_{\boldsymbol{x}_{k+1}} P(\boldsymbol{e}_{k+1} \mid \boldsymbol{x}_{k+1}) P(\boldsymbol{e}_{k+2:t} \mid \boldsymbol{x}_{k+1}) \boldsymbol{P}(\boldsymbol{x}_{k+1} \mid \boldsymbol{x}_k)$

Linear form

- $\boldsymbol{b}_{k+1:t} = \boldsymbol{P}(\boldsymbol{b}_{k+2:t}, \boldsymbol{e}_{k+1})$
- $\boldsymbol{b}_{k+1:t} = \boldsymbol{T}\boldsymbol{O}_{k+1} \boldsymbol{b}_{k+2:t}$

Forward-Backward Algorithm

- $\boldsymbol{f}_{1:k} = \alpha \textsc{Forward}(\boldsymbol{f}_{1:k-1}, \boldsymbol{e}_k)$
- $\boldsymbol{b}_{k+1:t} = \alpha \textsc{Backward}(\boldsymbol{b}_{k+2:t}, \boldsymbol{e}_{k+1})$
- $\boldsymbol{P}(\boldsymbol{X}_k \mid \boldsymbol{e}_{1:t}) = \alpha \boldsymbol{f}_{1:k} \times \boldsymbol{b}_{k+1:t}$

Viterbi Algorithm

- Predict $\{x_1, \cdots, x_t\}$ based on the observations $\{e_1, \cdots, e_t\}$
- $$\arg\max_{x_1, \cdots, x_t} P(x_1, \cdots, x_t \mid e_1, \cdots, e_t)$$
- MAP (Maximize A Posteriori)

Viterbi Algorithm

a recursive relationship between most likely paths to each state $\boldsymbol{x}_{t+1}$ and most likely paths to each state $\boldsymbol{x}_t$

$$
\begin{aligned}
&\max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_t} \boldsymbol{P}(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_t,\boldsymbol{X}_{t+1} \mid \boldsymbol{e}_{1:t+1}) \\
&= \max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_t} (\alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{x}_t) P(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})) \\
&= \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_t} (\boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{x}_t) P(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})) \\
&= \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \\
&\max_{\boldsymbol{x}_t} \left[\max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1}} (\boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{x}_t) P(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})) \right] \\
&= \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{X}_{t+1}) \\
&\max_{\boldsymbol{x}_t} \left[\boldsymbol{P}(\boldsymbol{X}_{t+1} \mid \boldsymbol{x}_t) \max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1}} (P(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})) \right]
\end{aligned}
$$

Viterbi Algorithm

- Let $\boldsymbol{m}_{1:t}$ be $\max_{\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1}} P(\boldsymbol{x}_1,\cdots,\boldsymbol{x}_{t-1},\boldsymbol{X}_t \mid \boldsymbol{e}_{1:t})$
- “Message” propagation
- $\boldsymbol{m}_{1:t+1} = \alpha \textsc{Viterbi}(\boldsymbol{m}_{1:t}, \boldsymbol{e}_{t+1})$

Viterbi Algorithm

Algorithm

$\boldsymbol{m}_{1:t} = \max_{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{t-1}} P(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_{t-1}, \boldsymbol{x}_t \mid \boldsymbol{e}_{1:t})$

- Initialization $\boldsymbol{m}_{1:t} = \boldsymbol{P}(\boldsymbol{e}_1 \mid \boldsymbol{x}_1) P(\boldsymbol{x}_1)$
- Indunction
  - $\boldsymbol{m}_{1:t+1} = \alpha \boldsymbol{P}(\boldsymbol{e}_{t+1} \mid \boldsymbol{x}_{t+1}) \max_{\boldsymbol{x}_t} [\boldsymbol{P}(\boldsymbol{x}_{t+1} \mid \boldsymbol{x}_t) \boldsymbol{m}_{1:t}]$
  - $x^*_{t+1} = \arg\max_{x_{t+1}} \boldsymbol{m}_{1:t+1}$