1 Introduction

In this report, we want to introduce a method to find the most-probable history of a Markov chain, that is, to find the path which has highest probability.

First, we discuss the problem in a non-hidden Markov chain. And later, the same problem in a hidden Markov chain will be covered. The solution is well-known as the Viterbi algorithm.

To make the problem simple, we discuss in a Markov chain which has only finite number of states.

2 Non-Hidden Cases

Let $(\Omega, \mathcal{F}, \mathcal{P})$ be a probability space.
Suppose that we have a Markov chain $(X_n)_{n\geq 0}$ with finite state-space and the transition matrix P.
Since the state space is finite, there must be an one-to-one correspondence between the state space and $I = \{1, 2, \dots, N\}$.
Thus it is reasonable to consider $X_n : \Omega \to I$ for all $n \in \mathbb{N}_0$, and $P = p[i, j]_{i,j\in I}$.

Here are some notations we use:

- $\mathbb{N}$ denotes the set $\{1, 2, 3, \dots\}$
- $\mathbb{N}_0$ denotes the set $\{0\} \cup \mathbb{N}$
- $x^{(n)} = (x_0, x_1, \dots, x^n) \in I^{n+1}$ is an apparent variable denoting a state sequence of length n for some $n \in \mathbb{N}_0$
- $\mathbf{X}^{(n)} = (X_0, X_1, \dots, X_n)$ denotes a Markov history from time 0 to some $n \in \mathbb{N}_0$

Our problem is described as follow:
Given an initial distribution $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_N)$ and a termination time $T \in \mathbb{N}$, what is the deterministic state sequence $x^* = (x^*_0, x^*_1, x^*_2, \dots, x^*_T)$ satisfying

$$\mathcal{P}(\mathbf{X}^{(T)} = x^*) = \max\left\{\mathcal{P}(\mathbf{X}^{(T)} = x^{(T)}) \mid x^{(T)} \in I^{T+1}\right\} \tag{1}$$

Assume that such x^* is unique, the representation of equation (1) is equivalent to

$$x^* = \arg\max_{x^{(T)}} \left\{\mathcal{P}(\mathbf{X}^{(T)} = x^{(T)})\right\} \tag{2}$$

Now we want to show that the sequence x^* can be found recursively.
First, we know the (a priori) probability of a deterministic path can be computed by

$$\begin{aligned}
\mathcal{P}(\mathbf{X}^{(n)} = x^{(n)}) &= \mathcal{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n) \\
&= \mathcal{P}(X_0 = x_0)\mathcal{P}(X_1 = x_1 \mid X_0 = x_0) \cdots \mathcal{P}(X_n = x_n \mid X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \\
&= \mathcal{P}(X_0 = x_0)\mathcal{P}(X_1 = x_1 \mid X_0 = x_0) \cdots \mathcal{P}(X_n = x_n \mid X_{n-1} = x_{n-1}) \quad \text{(by Markov assumption)} \\
&= \mathcal{P}(X_0 = x_0)\, p[x_0, x_1] \, \cdots \, p[x_{n-1}, x_n]
\end{aligned}$$

Before we move forward, we need a convenient notation.
For $n \in \mathbb{N}_0$ and $i \in I$, define $y(n, i)$ and $a(n, i)$.

- $y(n, i) = (y_0, y_1, \dots, y_{n-1}, i) \in I^{n+1}$ denotes the state sequence satisfies

$$\mathcal{P}(\mathbf{X}^{(n)} = y(n, i)) = \max\left\{\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i) \mid x^{(n-1)} \in I^n\right\}$$

- $a(n, i)$ denotes the probability $\max\left\{\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i) \mid x^{(n-1)} \in I^n\right\}$

Given $n \in \mathbb{N}_0$ and $i \in I$, since the state-space is finite, $y(n, i)$ is guarantee to exist.
Though $y(n, i)$ may not be unique, all of them would have the same probability, that is, $a(n, i)$ is unique. Therefore, we can assign any sequence $x^{(n)} \in I^{n+1}$ with $x_n = i$ satisfying $\mathcal{P}(\mathbf{X}^{(n)} = x^{(n)}) = a(n, i)$ as $y(n, i)$.

Based on the knowing of $y(n, j)\ \forall\, j \in I$, we want to obtain $y(n + 1, i)$.
We have

$$\mathcal{P}(X_0 = x_0, X_1 = x_1, \dots, X_{n+1} = x_{n+1}) = \mathcal{P}(X_0 = x_0, X_1 = x_1, \dots, X_n = x_n)\, p[x_n, x_{n+1}] \tag{3}$$

For $i \in I$, by definition, $y(n+1,i)$ satisfies

$$\mathcal{P}(y(n+1,i)) = \max\left\{\mathcal{P}(\mathbf{X}^{(n+1)} = x^{(n+1)}) \mid x^{(n+1)} = (x_0, x_1, \ldots, x_n, i)\right\} \tag{4}$$

Since every path 123

$$\begin{aligned}
\mathcal{P}(y(n+1,i)) &= \max\left\{\mathcal{P}(\mathbf{X}^{(n)} = x^{(n)}, X_{n+1} = i) \mid x^{(n+1)} \in I^{n+1}\right\} \\
&= \max\left\{\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, X_{n+1} = i) \mid x^{(n-1)} \in I^n, j \in I\right\} \qquad \text{by (3)} \\
&= \max\left\{\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j)\mathcal{P}(X_{n+1} = i \mid X_n = j) \mid x^{(n-1)} \in I^n, j \in I\right\} \qquad \text{(by Markov assumption)} \\
&= \max\left\{\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j)\, p[j,i] \mid x^{(n-1)} \in I^n, j \in I\right\} \qquad (5)
\end{aligned}$$

Since we already know the sequence to maximize $\mathcal{P}(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j)$ for each $j \in I$, we do not need to go through all I^{n+1} possible cases. We can search among the cases conditioning on the last state n. Hence,

$$\begin{aligned}
\mathcal{P}(y(n+1,i)) &= (5) \\
&= \max\left\{\mathcal{P}(\mathbf{X}^{(n)} = y(n,j))\, p[j,i] \mid j \in I\right\} \qquad (6)
\end{aligned}$$

Now, we need a starting point to complete the recursive algorithm.
Since we have the initial distribution of the chain, the algorithm is complete, as follow:

Initialization:
 for $i \in I$ **do**
  $y(0,i) \leftarrow \lambda_i$
 end for
Recursion:
 for $n = 1$ to $n = T$ **do**
  for $i \in I$ **do**
   $k \leftarrow \arg\max_j \left\{\mathcal{P}(\mathbf{X}^{(n)} = y(n,j))\, p[j,i] \mid j \in I\right\}$
   $y(n,i) \leftarrow (y(n-1,k), i)$
  end for
 end for
Termination:
 $x^* \leftarrow \arg\max_{y(T,i)} \left\{\mathcal{P}(\mathbf{X}^{(T)} = y(T,i)) \mid i \in I\right\}$

---

3 Hidden Cases

In a hidden Markov chain, we can never know the state sequence even when the chain has terminated.
Hence there is a small difference between the problem we are discussing in these two situations.

- In the non-hidden cases, we try to predict the most probable result before we would see it.
- In hidden cases, we try to figure out what state sequence has gone through after we have seen some outcome.