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 $I = \{1, 2, \ldots, N\}$.

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 $\{1, 2, 3, \ldots\}$

- $\mathbb{N}_0$ denotes the set $\{0\} \cup \mathbb{N}$

- $x^{(n)} = (x_0, x_1, \ldots, x^n) \in I^{n+1}$ is an apparent variable denoting a state sequence of length n for some $n \in \mathbb{N}_0$

- $\mathbf{X}^{(n)} = (X_0, X_1, \ldots, X_n)$ denotes a Markov history from time 0 to some $n \in \mathbb{N}_0$

Our problem is described as follow:

Given an initial distribution $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ and a termination time $T \in \mathbb{N}$, what is the deterministic state sequence $x^* = (x_0^*, x_1^*, x_2^*, \ldots, x_T^*)$ 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

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

Before we move forward, we need a convenient notation.

For $n \in \mathbb{N}_0$ and $i \in I$, define $y(n, i)$ and $a(n, i)$.

- $y(n, i) = (y_0, y_1, \ldots, y_{n-1}, i) \in I^{n+1}$ 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

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

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

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

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

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,

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

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**
        $a(0, i) \leftarrow \lambda_i$
        $y(0, i) \leftarrow i$
    end for
Recursion:
    for $n = 1$ to $n = T$ **do**
        for $i \in I$ **do**
            $a(n, i) \leftarrow \max_{j} \left\{ a(n-1, j) \, p[j, i] \mid j \in I \right\}$
            $k \leftarrow \arg\max_{j} \left\{ a(n-1, j) \, p[j, i] \mid j \in I \right\}$
            $y(n, i) \leftarrow (y(n-1, k), i)$
        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 was gone through after we have seen some outcome.

Most of the variables remain the same as before. Here, we have additional information, a sequence of observations $O = (o_1, o_2, \ldots, o_T)$, in which each element comes from the observation space $Q = \{1, 2, 3, \ldots, V\}$ where $V \in \mathbb{N}$.
Though the state sequence is hidden, we have an observation likelihood (probability) functions b such that for $1 \leq n \leq T$ and $i \in I$, $b(n, i) = P(o_n \mid X_n = i)$.

Then our problem is described as follow:

Given an initial distribution $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)$, the termination time $T \in \mathbb{N}$, and a sequence of oberservations $O = (o_1, o_2, \ldots, o_T)$, what is the deterministic sequence $x^* = (x_0^*, x_1^*, x_2^*, \ldots, x_T^*)$ satisfying

$$\mathcal{P}(\mathbf{X}^{(T)} = x^*) = \max\left\{ \mathcal{P}\left(\mathbf{X}^{(T)} = x^{(T)}, O\right) \Big| x^{(T)} \in I^{T+1}\right\} \tag{7}$$

And there are two assumptions:

- $o_1, o_2, \ldots, o_T$ are independent

- At any time step, the observation depends only on the state at that time

Similarly as in the previous section, for $1 \le n \le T$, define $O^{(n)}, y(n, i), \alpha(n, i)$.

- $O^{(n)} = (o_1, o_2, \ldots, o_n)$ denotes the observation sequence to time n

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

$$\mathcal{P}\left(\mathbf{X}^{(n)} = y(n, i) \mid O^{(n)}\right) = \max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i \mid O^{(n)}\right) \Big| x^{(n-1)} \in I^n\right\} \tag{8}$$

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

Since

$$\mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i \mid O^{(n)}\right) = \frac{\mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i, O^{(n)}\right)}{\mathcal{P}(O^{(n)})} \tag{9}$$

then

$$\arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i \mid O^{(n)}\right) \Big| x^{(n-1)} \in I^n\right\}$$
$$= \arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = i, O^{(n)}\right) \Big| x^{(n-1)} \in I^n\right\} \tag{10}$$

Suppose that we know $y(n, j)\ \forall\, j \in I$.
We want to obtain $y(n+1, i) = (y_0, y_1, \ldots, y_n, i)$. According to equation (10),

$$(y_0, y_1, \ldots, y_n) = \arg\max_{x^{(n)}}\left\{ \mathcal{P}\left(\mathbf{X}^{(n)} = x^{(n)}, X_{n+1} = i, O^{(n+1)}\right) \Big| x^{(n)} \in I^{n+1}\right\}$$
$$= \arg\max_{(x^{(n-1)}, j)}\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, X_{n+1} = i, O^{(n+1)}\right) \Big| x^{(n-1)} \in I^n, j \in I\right\}$$
$$= \arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, X_{n+1} = i, O^{(n)}, o_{n+1}\right) \Big| x^{(n-1)} \in I^n, j \in I\right\} \tag{11}$$

Since o_{n+1} is independent with $O^{(n)}$ and depends only on X_{n+1},

$$(11) = \arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, X_{n+1} = i, O^{(n)}\right) \mathcal{P}\left(o_{n+1} \mid X_{n+1} = i\right) \Big| x^{(n-1)} \in I^n, j \in I\right\} \tag{12}$$
$$= \arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, X_{n+1} = i, O^{(n)}\right) b(n+1, i) \Big| x^{(n-1)} \in I^n, j \in I\right\} \tag{13}$$
$$= \arg\max\left\{ \mathcal{P}\left(\mathbf{X}^{(n-1)} = x^{(n-1)}, X_n = j, O^{(n)}\right) p[j, i]\, b(n+1, i) \Big| x^{(n-1)} \in I^n, j \in I\right\} \quad \text{(by Markov assumption)} \tag{14}$$

Similar to what we have done in (5), we have the result

$$(y_0, y_1, \ldots, y_n) = (14) \tag{15}$$
$$= \left(y(n-1, j),\ \arg\max_j\left\{ \mathcal{P}\left(\mathbf{X}^{(n)} = y(n, j)\right) p[j.i]\, b(n+1, i) \Big| j \in I\right\} \right) \tag{16}$$

Hence in the hidden case, we can solve the problem recursively as well.

Here is the pseudocode of the algorithm:

Initialization:

    for $i \in I$ **do**

        $\alpha(0, i) \leftarrow \lambda_i$

        $y(0, i) \leftarrow i$

    end for

Recursion:

    for $n = 1$ **to** $n = T$ **do**

        for $i \in I$ **do**

            $\alpha(n, i) \leftarrow \max_{j} \{\alpha(n - 1, j)\, b(n, i)\, p[j, i] \mid j \in I\}$

            $k \leftarrow \arg\max_{j} \{\alpha(n - 1, j)\, b(n, i)\, p[j, i] \mid j \in I\}$

            $y(n, i) \leftarrow (y(n - 1, k), i)$

        end for

    end for

Termination:

$x^* \leftarrow \arg\max_{y(T, i)} \left\{ \mathcal{P}\left(\mathbf{X}^{(T)} = y(T, i)\right) b(T, i) \mid i \in I \right\}$