

Hidden Markov Model

Depu Meng

Oct. 2018

1 Basic Concepts & Examples

1.1 Basic Concepts

Definition 1.1 Assume $X = \{X_n; n \geq 1\}$ is a Markov chain over finite state space $\Phi = \{1, 2, \dots, K\}$, if states of X are unobservable, $Y = \{Y_n; n \geq 1\}$ is an observable random variable series over finite set $V = \{v_1, v_2, \dots, v_L\}$ that is correlated with X , then $(X, Y) = \{(X_n, Y_n); n \geq 1\}$ is a *Hidden Markov Chain*.

Denote $\pi = \{\pi_1, \pi_2, \dots, \pi_k\}$ is the initial distribution of X , $A = [a_{ij}]$ is the transition matrix of X

$$a_{ij} = P\{X_{n+1} = j | X_n = i\}, i, j \in \Phi \quad (1)$$

is the one-step transition probability of X . Denote

$$b_{ij} = P\{Y_n = v_j | X_n = i\}, i \in \Phi, v_j \in V \quad (2)$$

represents the probability that Y equals to v_j given X is at state i at time n . If X is homogeneous, then Y is also irrelevant with time n . Denote $B = [b_{ij}]$ as the observation probability matrix. Due to the unobservability of X , π, A, B can not be directly measured. Generally, we call parameter set $\lambda = \{\pi, A, B\}$ the math model of Hidden Markov chain (X, Y) , as well as *Hidden Markov Model* (HMM).

Property 1.1 Assume N is the length of time on observation, denote $\mathbf{X} = \{X_1, X_2, \dots, X_N\}$, $\mathbf{Y} = \{Y_1, Y_2, \dots, Y_N\}$ as the sample series of Markov chain X and observation series Y in the time period $1 \sim N$ respectively, then the joint distribution of \mathbf{X} and \mathbf{Y} satisfy the following *Hidden Markov Condition*.

$$P\{\mathbf{X} = x, \mathbf{Y} = y\} = \pi_{i_1} b_{i_1 j_1} \dots b_{i_{N-1} j_{N-1}} b_{i_N j_N} \quad (3)$$

where $x = \{i_1, \dots, i_N\}$, $y = \{v_{j_1}, \dots, v_{j_N}\}$.

1.2 Examples

Coin flipping problem Stochastically choose a coin from two coins indexed 1 and 2, and then toss it, observe the result and repeat this process. Denote X_n as the coin chosen in n^{th} time, then $X = \{X_n; n \geq 1\}$ is a Markov chain over state space $\Phi = \{1, 2\}$. Denote Y_n as the result of n^{th} experiment, then $Y = \{Y_n; n \geq 1\}$

is observation series that takes values from $V = \{H, T\}$, where H means the front side and T means the back side. State transition matrix and observation probability matrix are

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \quad (4)$$

where summation of each row equals to 1. For initial distribution $\pi = \{\pi_1, \pi_2\}$, we have $\pi_1 + \pi_2 = 1$. Then $\lambda = \{\pi, A, B\}$ is the corresponding HMM of Hidden Markov chain (X, Y) , which contains 5 unknown parameters.

2 Basic Problems & Solutions

2.1 Basic Problems

Problem 1 For a specific observation sample series $\mathbf{Y} = \{Y_1, \dots, Y_n\}$, known it is generated by one of the given HMMs, determine which model that generates the sample series is called *Pattern Recognition* problem, as well as *classification* problem.

Problem 2 From a series of observation samples $\mathbf{Y} = \{Y_1, \dots, Y_n\}$ and known HMM $\lambda = \{\pi, A, B\}$, give the best estimate of hidden states is called *State Estimate* problem, as well as *Decoding* problem.

Problem 3 From a series of observation samples $\mathbf{Y} = \{Y_1, \dots, Y_n\}$, give the best estimate of parameter set $\lambda = \{\pi, A, B\}$, we call it *Model Learning* problem.

2.2 Solutions

Solution to Problem 1 To solve Problem 1 is mainly to solve $P\{\mathbf{Y}|\lambda\}$, for different λ , we can compare with their corresponding $P\{\mathbf{Y}|\lambda\}$, and choose the largest one. So how to compute $P\{\mathbf{Y}|\lambda\}$?

$$P\{\mathbf{Y} = y|\lambda\} = \sum_x P\{\mathbf{Y} = y, \mathbf{X} = x|\lambda\} \quad (5)$$

$$= \sum_x \pi_{i_1} b_{i_1 j_1} \dots b_{i_{N-1} j_{N-1}} b_{i_N j_N} \quad (6)$$

In order to make the writing more convenient, we introduce notation $\alpha_n(i)$ and $\beta_n(i)$ first.

$$\alpha_n(i) = P\{Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = i|\lambda\} \quad (7)$$

$$\beta_n(i) = P\{Y_{n+1} = v_{j_{n+1}}, \dots, Y_N = v_{j_N} | X_n = i, \lambda\} \quad (8)$$

Recurrence relation of $\alpha_n(i)$.

$$\alpha_n(i) = \sum_{j=1}^K \alpha_n(j) a_{ji} b_{ij_{n+1}}, n = 1, \dots, N-1 \quad (9)$$

proof

$$\alpha_{n+1}(i) = P\{Y_1 = v_{j_1}, Y_n = v_{j_n}, Y_{n+1} = v_{j_{n+1}}, X_{n+1} = i | \lambda\} \quad (10)$$

$$= \sum_{k=1}^K P\{Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, Y_{n+1} = v_{j_{n+1}}, X_n = k, X_{n+1} = i | \lambda\} \quad (11)$$

$$= \sum_{k=1}^K P_A P_B P_C \quad (12)$$

where

$$P_A = P\{Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k | \lambda\} \quad (13)$$

$$P_B = P\{X_{n+1} = i | Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k, \lambda\} \quad (14)$$

$$P_C = P\{Y_{n+1} = v_{j_{n+1}} | Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k, X_{n+1} = i, \lambda\} \quad (15)$$

Notice that for P_A , we have

$$P_A = P\{Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k | \lambda\} = \alpha_n(k) \quad (16)$$

$$(17)$$

For P_B , we have

$$P_B = P\{X_{n+1} = i | Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k, \lambda\} \quad (18)$$

$$= P\{X_{n+1} = i | X_n = k, \lambda\} \quad (19)$$

$$= a_{ki} \quad (20)$$

For P_C , we have

$$P_C = P\{Y_{n+1} = v_{j_{n+1}} | Y_1 = v_{j_1}, \dots, Y_n = v_{j_n}, X_n = k, X_{n+1} = i, \lambda\} \quad (21)$$

$$= P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = i, \lambda\} \quad (22)$$

$$= b_{ij_{n+1}} \quad (23)$$

Combine the three equations together, we have

$$\alpha_{n+1}(i) = \sum_{k=1}^K P_A P_B P_C \quad (24)$$

$$= \sum_{k=1}^K \alpha_n(k) a_{ki} b_{ij_{n+1}} \quad (25)$$

□

Recurrence relation of $\beta_n(i)$

$$\beta_n(i) = \sum_{j=1}^K \beta_{n+1}(j) a_{ij} b_{jj_{n+1}}, n = 1, \dots, N-1 \quad (26)$$

proof

$$\beta_n(i) = P\{Y_{n+1} = v_{j_{n+1}}, \dots, Y_N = v_{j_N} | X_n = i, \lambda\} \quad (27)$$

$$= \sum_{k=1}^K P\{Y_{n+1} = v_{j_{n+1}}, \dots, Y_N = v_{j_N}, X_{n+1} = k | X_n = i, \lambda\} \quad (28)$$

$$= \sum_{k=1}^K P_X P_Y P_Z \quad (29)$$

where

$$P_X = P\{X_{n+1} = k | X_n = i, \lambda\} \quad (30)$$

$$P_Y = P\{Y_{n+2} = v_{j_{n+1}}, \dots, Y_N = v_{j_N} | X_{n+1} = k, X_n = i, \lambda\} \quad (31)$$

$$P_Z = P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, X_n = i, Y_{n+2} = v_{j_{n+1}}, \dots, Y_N = v_{j_N}, \lambda\} \quad (32)$$

For P_X , we have

$$P_X = P\{X_{n+1} = k | X_n = i, \lambda\} = a_{ik} \quad (33)$$

For P_Y , we have

$$P_Y = P\{Y_{n+2} = v_{j_{n+1}}, \dots, Y_N = v_{j_N} | X_{n+1} = k, X_n = i, \lambda\} \quad (34)$$

$$= P\{Y_{n+2} = v_{j_{n+1}}, \dots, Y_N = v_{j_N} | X_{n+1} = k, \lambda\} \quad (35)$$

$$= \beta_{n+1}(k) \quad (36)$$

For P_Z , we have

$$P_Z = P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, X_n = i, Y_{n+2} = v_{j_{n+1}}, \dots, Y_N = v_{j_N}, \lambda\} \quad (37)$$

$$= P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, \lambda\} \quad (38)$$

$$= b_{kj_{n+1}} \quad (39)$$

Combine the three equations, we have

$$\beta_n(i) = \sum_{k=1}^K P_X P_Y P_Z \quad (40)$$

$$= \sum_{k=1}^K \beta_{n+1}(k) a_{ik} b_{kj_{n+1}} \quad (41)$$

□

From the $\alpha_n(i), \beta_n(i)$ defined above, we have

$$P\{\mathbf{Y} = y | \lambda\} = \sum_{j=1}^K \alpha_N(j) \quad (42)$$

$$P\{\mathbf{Y} = y | \lambda\} = \sum_{j=1}^K \beta_1(j) \pi_j b_{jj_1} \quad (43)$$

$$(44)$$

If we divide observation series $\mathbf{Y} = \{Y_1, \dots, Y_N\}$ into $\{Y_1, \dots, Y_n\}$ and $\{Y_{n+1}, \dots, Y_N\}$, then we have

$$P\{\mathbf{Y} = y, X_n = i | \lambda\} = \alpha_n(i) \beta_n(i) \quad (45)$$

Solution to Problem 2 For $n = 1, 2, \dots, N$, denote

$$\gamma_n(i) = P\{X_n = i | Y_1 = v_{j_1}, \dots, Y_N = v_{j_N}, \lambda\} \quad (46)$$

as the probability of state i given observation series $Y_1 = v_{j_1}, \dots, Y_N = v_{j_N}$ and model parameter set λ . For $\gamma_n(i)$, we have

$$\gamma_n(i) = \frac{P\{Y_1 = v_{j_1}, \dots, Y_N = v_{j_N}, X_n = i | \lambda\}}{P\{Y_1 = v_{j_1}, \dots, Y_N = v_{j_N} | \lambda\}} \quad (47)$$

$$= \frac{P\{Y_1 = v_{j_1}, \dots, Y_N = v_{j_N}, X_n = i | \lambda\}}{\sum_{k=1}^K P\{Y_1 = v_{j_1}, \dots, Y_N = v_{j_N}, X_n = k | \lambda\}} \quad (48)$$

$$= \frac{\alpha_n(i) \beta_n(i)}{\sum_{k=1}^K \alpha_n(k) \beta_n(k)} \quad (49)$$

As we can see, $\gamma_n(i)$ is a probability measure satisfying $\sum_{k=1}^K \gamma_n(k) = 1$. If

$$i^* = \arg \max_{1 \leq i \leq K} \gamma_n(i) \quad (50)$$

then we select $\hat{X}_n = i^*$ as the estimate of time n . However, this algorithm ignore the connection between different time, like if some $a_{ij} = 0$, some optimal series can not be reached.

Viterbi Algorithm Viterbi algorithm is a progressive optimization algorithm based on *Dynamic Programming*, denote

$$\delta_n(i) = \max_{i_1 i_2 \dots i_{n-1}} P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1; Y_n = v_{j_n}, \dots, Y_1 = v_{j_1} | \lambda\} \quad (51)$$

Why $\delta_n(i)$? Because initially we want to compute

$$\arg \max_{i_1 i_2 \dots i_{n-1} i} P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1 | Y_n = v_{j_n}, \dots, Y_1 = v_{j_1}, \lambda\} \quad (52)$$

we have

$$P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1 | Y_n = v_{j_n}, \dots, Y_1 = v_{j_1}, \lambda\} \quad (53)$$

$$= \frac{P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1; Y_n = v_{j_n}, \dots, Y_1 = v_{j_1} | \lambda\}}{\sum_{i_1} \sum_{i_2} \dots \sum_{i_i} P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1; Y_n = v_{j_n}, \dots, Y_1 = v_{j_1} | \lambda\}} \quad (54)$$

due to the denominator is irrelevant to i_1, i_2, \dots, i , so equation 52 is equals to

$$\arg \max_{i_1 i_2 \dots i_{n-1} i} P\{X_n = i, X_{n-1} = i_{n-1}, \dots, X_1 = i_1 | Y_n = v_{j_n}, \dots, Y_1 = v_{j_1}, \lambda\} \quad (55)$$