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

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

$$a_{ij} = P\{X_{n+1} = j | X_n = i\}, i, j \in \varPhi \tag{1}$$

is the one-step transition probality of X. Denote

$$b_{ij} = P\{Y_n = v_j | X_n = i\}, i \in \varPhi, v_j \in V \tag{2}$$

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 irrelavent 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 $\mathbf{X} = \{X_1, X_2, ..., X_N\}$, $\mathbf{Y} = \{Y_1, Y_2, ..., Y_N\}$ 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*.

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

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

1.2 Examples

Coin flipping problem Stochasticly 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 t^{th} time, then $X = \{X_n; n \geq 1\}$ is a Markov chain over state space $\varPhi = \{1, 2\}$. 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

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

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 $\mathbf{Y} = \{Y_1, ..., Y_n\}$, 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 $\mathbf{Y} = \{Y_1, ..., Y_n\}$ 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 $\mathbf{Y} = \{Y_1, ..., Y_n\}$, 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 $P\{\mathbf{Y}|\lambda\}$?

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

$$= \sum_x \pi_{i_1} b_{i_1 j_1} ... b_{i_{N-1} j_{N-1}} b_{i_{N-1} j_N} b_{i_N j_N} \tag{6}$$

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

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

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

Recurrence relation of $\alpha_n(i)$.

$$\alpha_n(i) = \sum_{j=1}^{K} \alpha_n(j) a_{ji} b_{i j_{n+1}}, n = 1, ..., N-1 \tag{9}$$

proof

$$\alpha_{n+1}(i) = P\{Y_1 = v_{j_1}, Y_n = v_{j_n}, Y_{n+1} = v_{j_{n+1}}, X_{n+1} = i|\lambda\} \tag{10}$$
$$= \sum_{k=1}^{K} P\{Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, Y_{n+1} = v_{j_{n+1}}, X_n = k, X_{n+1} = i|\lambda\} \tag{11}$$
$$= \sum_{k=1}^{K} P_A P_B P_C \tag{12}$$

where

$$P_A = P\{Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k|\lambda\} \tag{13}$$
$$P_B = P\{X_{n+1} = i|Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k, \lambda\} \tag{14}$$
$$P_C = P\{Y_{n+1} = v_{j_{n+1}}|Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k, X_{n+1} = i, \lambda\} \tag{15}$$

Notice that for P_A, we have

$$P_A = P\{Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k|\lambda\} = \alpha_n(k) \tag{16}$$
$$\tag{17}$$

For P_B, we have

$$P_B = P\{X_{n+1} = i|Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k, \lambda\} \tag{18}$$
$$= P\{X_{n+1} = i|X_n = k, \lambda\} \tag{19}$$
$$= a_{ki} \tag{20}$$

For P_C, we have

$$P_C = P\{Y_{n+1} = v_{j_{n+1}}|Y_1 = v_{j_1}, ..., Y_n = v_{j_n}, X_n = k, X_{n+1} = i, \lambda\} \tag{21}$$
$$= P\{Y_{n+1} = v_{j_{n+1}}|X_{n+1} = i, \lambda\} \tag{22}$$
$$= b_{ij_{n+1}} \tag{23}$$

Combine the three equations together, we have

$$\alpha_{n+1}(i) = \sum_{k=1}^{K} P_A P_B P_C \tag{24}$$
$$= \sum_{k=1}^{K} \alpha_n(k) a_{ki} b_{ij_{n+1}} \tag{25}$$

□

Recurrence relation of $\beta_n(i)$

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

proof

$$\beta_n(i) = P\{Y_{n+1} = v_{j_{n+1}}, ..., Y_N = v_{j_N} | X_n = i, \lambda\} \tag{27}$$
$$= \sum_{k=1}^{K} P\{Y_{n+1} = v_{j_{n+1}}, ..., Y_N = v_{j_N}, X_{n+1} = k | X_n = i, \lambda\} \tag{28}$$
$$= \sum_{k=1}^{K} P_X P_Y P_Z \tag{29}$$

where

$$P_X = P\{X_{n+1} = k | X_n = i, \lambda\} \tag{30}$$
$$P_Y = P\{Y_{n+2} = v_{j_{n+1}}, ..., Y_N = v_{j_n} | X_{n+1} = k, X_n = i, \lambda\} \tag{31}$$
$$P_Z = P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, X_n = i, Y_{n+2} = v_{j_{n+1}}, ..., Y_N = v_{j_n}, \lambda\} \tag{32}$$

For P_X, we have

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

For P_Y, we have

$$P_Y = P\{Y_{n+2} = v_{j_{n+1}}, ..., Y_N = v_{j_n} | X_{n+1} = k, X_n = i, \lambda\} \tag{34}$$
$$= P\{Y_{n+2} = v_{j_{n+1}}, ..., Y_N = v_{j_n} | X_{n+1} = k, \lambda\} \tag{35}$$
$$= \beta_{n+1}(k) \tag{36}$$

For P_Z, we have

$$P_Z = P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, X_n = i, Y_{n+2} = v_{j_{n+1}}, ..., Y_N = v_{j_n}, \lambda\} \tag{37}$$
$$= P\{Y_{n+1} = v_{j_{n+1}} | X_{n+1} = k, \lambda\} \tag{38}$$
$$= b_{kj_{n+1}} \tag{39}$$

Combine the three equations, we have

$$\beta_n(i) = \sum_{k=1}^{K} P_X P_Y P_Z \tag{40}$$
$$= \sum_{k=1}^{K} \beta_{n+1}(k) a_{ik} b_{kj_{n+1}} \tag{41}$$

□

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

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

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

$$\tag{44}$$

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

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

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

$$\gamma_n(i) = P\{X_n = i|Y_1 = v_{j_1}, ..., Y_N, \lambda\} \tag{46}$$

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

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

$$= \frac{P\{Y_1 = v_{j_1}, ..., Y_N = v_{j_N}, X_n = i|\lambda\}}{\sum_{k=1}^{K} P\{Y_1 = v_{j_1}, ..., Y_N = v_{j_N}, X_n = k|\lambda\}} \tag{48}$$

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

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

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

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

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

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

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

we have

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

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

due to the denominator is irrelavent to $i_1, i_2, ..., i$, so equtation 52 is equals to

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