

HMM模型--预测问题

在<https://github.com/Casey1203/ml-ease/blob/master/hmm/HMM%E6%A8%A1%E5%9E%8B-%E5%AE%9A%E4%B9%89.md>文档中，对HMM模型应用的三种问题，分为概率计算、参数学习、预测。在本文中对预测问题进行介绍。

4. 预测算法

4.1 近似算法

这个算法的思想比较简单，并且在概率计算部分，简单提及了。

回顾 $\gamma_t(i)$ 的定义，表示在时刻 t ，处于状态 i 的概率

$$\begin{aligned}\gamma_t(i) &= \frac{P(i_t = i, O|\lambda)}{\sum_{i=1}^N P(i_t = i, O|\lambda)} = P(i_t = i|O, \lambda) \\ &= \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)}\end{aligned}$$

因此，每个时刻，都有 N 个 γ ，选择可以使 γ 取到最大的状态作为最有可能的状态 i_t^* 。

$$i_t^* = \arg \max_{1 \leq i \leq N} \gamma_t(i), \quad t = 1, 2, \dots, T$$

从而可以得到最有可能的序列 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$

这种方法仅是对单独的每个时间点作预测，没有考虑前后状态的联系。因此，得到的序列 I^* 从整体上，不能保证是最有可能的状态序列。同时，由于不考虑整理，可能出现相邻的状态，转移概率为0。

4.2 维特比算法

维特比算法实际上是用动态规划，求解HMM的预测问题。

根据动态规划的原理，最优路径的特点：如果最优路径在时刻 t 通过结点 i_t^* ，那么整条路径 I^* 中，从结点 i_t^* 到结点 i_T^* 的部分 $I_{t \rightarrow T}^*$ ，相比于从 i_t^* 到 i_T^* 的所有可能的路径来说，一定是最优的。否则，如果存在一条从 i_t^* 到 i_T^* 的路径 $I_{t \rightarrow T}$ ，要好于当前找到的路径中结点 i_t^* 到结点 i_T^* 的部分 $I_{t \rightarrow T}^*$ ，那用 $I_{t \rightarrow T}$ 去替换 $I_{t \rightarrow T}^*$ ，可以得到一条比 I^* 更好的路径。

定义，在时刻 t ，状态为 i 的所有单个路径 (i_1, i_2, \dots, i_t) 中概率最大值为

$$\delta_t(i) = \max_{i_1, i_2, \dots, i_{t-1}} P(i_t = i, i_{t-1}, \dots, i_1, o_t, \dots, o_1 | \lambda), \quad i = 1, 2, \dots, N$$

这个max的意思是，在 t 时刻的状态已经确定为 i 的情况下，从总体上遍历 i_{t-1}, \dots, i_1 ，找到使得概率最大的路径 $I_{1 \rightarrow t-1}$ 。

因此 $\delta_t(i)$ 的递推公式为

$$\begin{aligned}\delta_{t+1}(i) &= \max_{i_1, i_2, \dots, i_t} P(i_{t+1} = i, i_t, \dots, i_1, o_{t+1}, \dots, o_1 | \lambda) \\ &= \max_{1 \leq j \leq N} [\delta_t(j) a_{ji}] b_i(o_{t+1}), \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T-1\end{aligned}$$

表示,时刻 t 位于某个状态 j ,并且在时刻 $t + 1$ 状态转移至 i 。外层 \max 选择 $t + 1$ 时刻位于状态 j ,内层循环要保证在第 t 时刻确定了状态为 j 的情况下,路径 $I_{1 \rightarrow t}$ 是概率最大的。完成了两层 \max 之后,确定了路径,还要再考虑观测概率 $b_i(o_{t+1})$,这就与路径无关了。

同时定义

$$\psi_t(i) = \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}], \quad i = 1, 2, \dots, N$$

表示,时刻 t 位于状态 i 的最优路径中,第 $t - 1$ 时刻(倒数第二个结点)位于 j 的状态。因此是对 $1 \leq j \leq N$ 求 $\arg \max$ 。

因此维特比算法的框架是

初值

$$\begin{aligned} \delta_1(i) &= \pi_i b_i(o_1), \quad i = 1, 2, \dots, N \\ \psi_1(i) &= 0, \quad i = 1, 2, \dots, N \end{aligned}$$

因为只包含一个结点,没有路径,所以 δ_1 不存在 \max 的问题。同时,定义 $t - 1 = 0$ 时刻的结点 ψ_1 为0。

递推,对于 $t = 2, 3, \dots, T$

$$\begin{aligned} \delta_t(i) &= \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}] b_i(o_t), \quad i = 1, 2, \dots, N; t = 1, 2, \dots, T - 1 \\ \psi_t(i) &= \arg \max_{1 \leq j \leq N} [\delta_{t-1}(j) a_{ji}], \quad i = 1, 2, \dots, N \end{aligned}$$

终止

根据递推过程,我们求得了 $\delta_T(i)$,表示到 T 时刻,状态为 i ,最优的路径对应的概率。由于我们不关心在 T 时刻具体位于哪个状态,因此

$$P^* = \max_{1 \leq i \leq N} \delta_T(i)$$

表示给定了观测序列 O 和模型参数 λ ,最优路径出现的概率。

同时,最优路径的最后一个结点,即 T 时刻的状态为

$$i_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)]$$

根据 ψ 的定义,表示在最优路径下,倒数第二个结点的状态,因此对 $t = T - 1, T - 2, \dots, 1$

$$i_t^* = \psi_{t+1}(i_{t+1}^*)$$

从而得到了最优的路径 $I^* = (i_1^*, i_2^*, \dots, i_T^*)$

下面给一个例子来运用维特比算法

给定模型参数 $\lambda = (A, B, \pi)$

$$A = \begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}, \quad B = \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix}, \quad \pi = (0.2, 0.4, 0.4)^T$$

已知观测序列 $O = (\text{红}, \text{白}, \text{红})$,求最优状态序列。

初始化,在 $t = 1$ 时,对每个状态 $i = 1, 2, 3$,求状态 i 观测到 $o_1 = \text{红}$ 的概率。

$$\delta_1(i) = \pi_i b_i(o_1) = \pi_i b_i(\text{红}), \quad i = 1, 2, 3$$

根据初始概率和观测概率计算

$$\delta_1(1) = 0.2 \times 0.5 = 0.10$$

$$\delta_1(2) = 0.4 \times 0.4 = 0.16$$

$$\delta_1(3) = 0.4 \times 0.7 = 0.28$$

同时, $\psi_1(i) = 0, \quad i = 1, 2, 3$

在 $t = 2$ 时刻, 对每个状态 $i, i = 1, 2, 3$, 求在 $t = 1$ 时刻状态为 j 观测为红, 并且在 $t = 2$ 时刻状态为 i 观测 $o_2 = \text{白}$ 的路径的最大概率, 记作 $\delta_2(i)$, 即

$$\delta_2(i) = \max_{1 \leq j \leq 3} [\delta_1(j) a_{ji}] b_i(o_2)$$

同时, 前一个结点的状态 j 为

$$\psi_2(i) = \arg \max_{1 \leq j \leq 3} [\delta_1(j) a_{ji}], \quad i = 1, 2, 3$$

代入模型参数计算, 计算在 $t = 2$ 时刻, 状态位于1的最优路径的概率

$$\begin{aligned} \delta_2(1) &= \max_{1 \leq j \leq 3} [\delta_1(j) a_{j1}] b_1(o_2) \\ &= \max_j \{0.1 \times 0.5, 0.16 \times 0.3, 0.28 \times 0.2\} \times 0.5 \\ &= 0.028 \end{aligned}$$

看到在计算 $\delta_2(1)$ 的 \max 时, 选择了状态3作为该结点的状态, 即

$$\psi_2(1) = 3$$

类似的, 可以计算在 $t = 2$ 时刻, 状态位于2或3的情况下的最优路径的概率

$$\begin{aligned} \delta_2(2) &= \max_{1 \leq j \leq 3} [\delta_1(j) a_{j2}] b_2(o_2) \\ &= \max_j \{0.1 \times 0.2, 0.16 \times 0.5, 0.28 \times 0.3\} \times 0.6 \\ &= 0.0504 \end{aligned}$$

$$\psi_2(2) = 3$$

$$\begin{aligned} \delta_2(3) &= \max_{1 \leq j \leq 3} [\delta_1(j) a_{j3}] b_3(o_2) \\ &= \max_j \{0.1 \times 0.3, 0.16 \times 0.2, 0.28 \times 0.5\} \times 0.3 \\ &= 0.042 \end{aligned}$$

$$\psi_2(3) = 3$$

计算 $t = 3$ 时,

$$\delta_3(i) = \max_{1 \leq j \leq 3} [\delta_2(j) a_{ji}] b_i(o_3), \quad \psi_3(i) = \arg \max_{1 \leq j \leq 3} [\delta_2(j) a_{ji}]$$

代入模型参数计算, 在 $t = 3$ 时刻, 状态分别位于1, 2, 3的最优路径的概率

$$\delta_3(1) = 0.00756, \quad \psi_3(1) = 2$$

$$\delta_3(2) = 0.01008, \quad \psi_3(2) = 2$$

$$\delta_3(3) = 0.0147, \quad \psi_3(3) = 3$$

因此最优路径的概率为

$$P^* = \max_{1 \leq i \leq 3} \delta_3(i) = \delta_3(3) = 0.0147$$

利用 ψ 回溯路径

在 $t = 3$ 时, $i_3^* = \arg \max_i [\delta_3(i)] = 3$

在 $t = 2$ 时, $i_2^* = \psi_3(i_3^*) = \psi_3(3) = 3$

在 $t = 1$ 时, $i_1^* = \psi_2(i_2^*) = \psi_2(3) = 3$

因此最优状态序列为 $I^* = (i_1^*, i_2^*, i_3^*) = (3, 3, 3)$

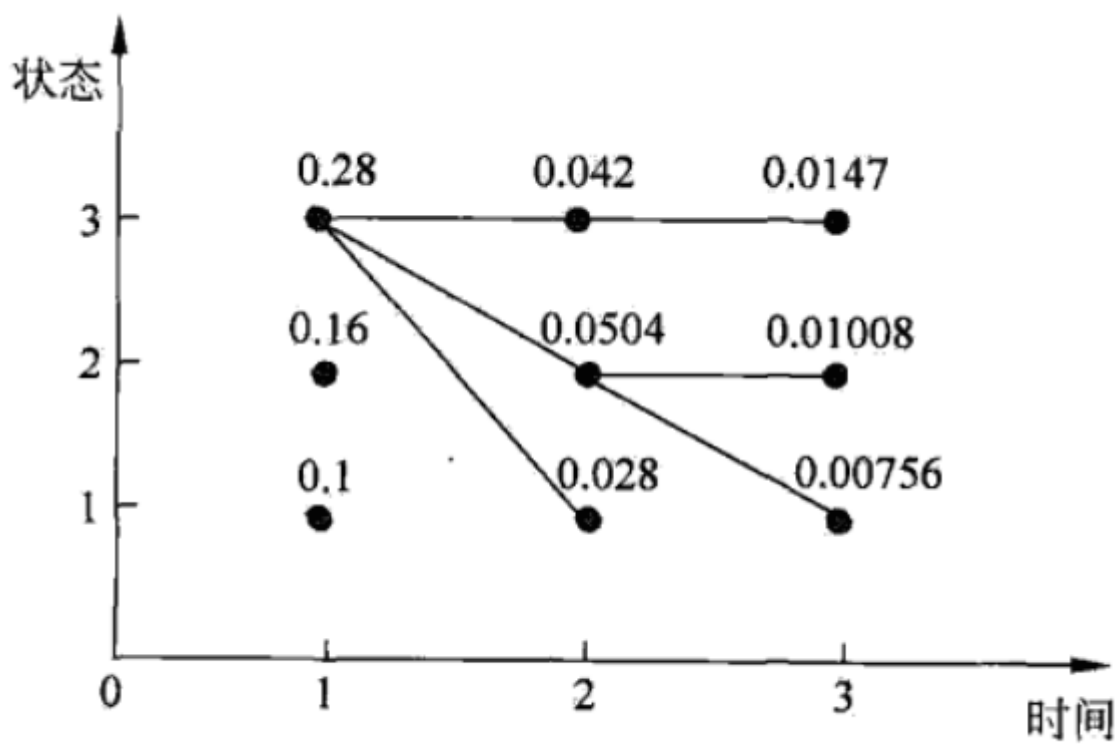


图 10.4 求最优路径