

Hmm 算法

hmm 三个问题

1. 给定参数 $A B P_i$, 求一个观察到显状态序列的概率
2. 解码, 给定参数 $A B P_i$, 求观察序列背后的隐状态序列
3. 求参数 $A B P_i$ 其中 A 是隐状态之间转移概率, B 是发射概率, 隐状态发射到显状态到概率, P_i 是初始隐状态概率

hmm 的假设:

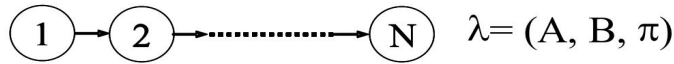
1. 一阶马尔科夫假设。隐状态只跟前一状态有关, 跟其他状态无关
2. 观测状态只与当前的隐状态有关, 跟其他状态无关

hmm 总结:

hmm 这个算法是动态规划和 em 算法+大量的公式证明组成, 其中大量的公式证明可以通过转化为路径问题来更加清晰的定位规律, 因为一大堆公式看的人头皮发麻, 往往找不到重点。公式隔两天就忘, 但是思想很容易记住。知道思想就够了。

第一个问题: 求观察到状态概率

全概率公式, 会造成时间复杂度 $O(n^t)$ 问题。原因是每一个隐状态都会有产生特定显状态的概率, 并且转移概率



$\bar{O} = o_1 o_2 o_3 \dots o_t \dots o_T$ observation sequence

$\bar{q} = q_1 q_2 q_3 \dots q_t \dots q_T$ state sequence

- **Problem 1:** Given λ and \bar{O} ,
find $P(\bar{O}|\lambda) = \text{Prob}[\text{observing } \bar{O} \text{ given } \lambda]$

- **Direct Evaluation:** considering all possible state sequence \bar{q}

$$P(\bar{O}|\lambda) = \sum_{\text{all } \bar{q}} P(\bar{O}, \bar{q}|\lambda) = \sum_{\text{all } \bar{q}} P(\bar{O}|\bar{q}, \lambda) P(\bar{q}|\lambda)$$

$$P(\bar{O}|\bar{q}, \lambda)$$

$$P(\bar{O}|\lambda) = \sum_{\text{all } \bar{q}} ([b_{q_1}(o_1) \cdot b_{q_2}(o_2) \cdot \dots \cdot b_{q_T}(o_T)] \cdot$$

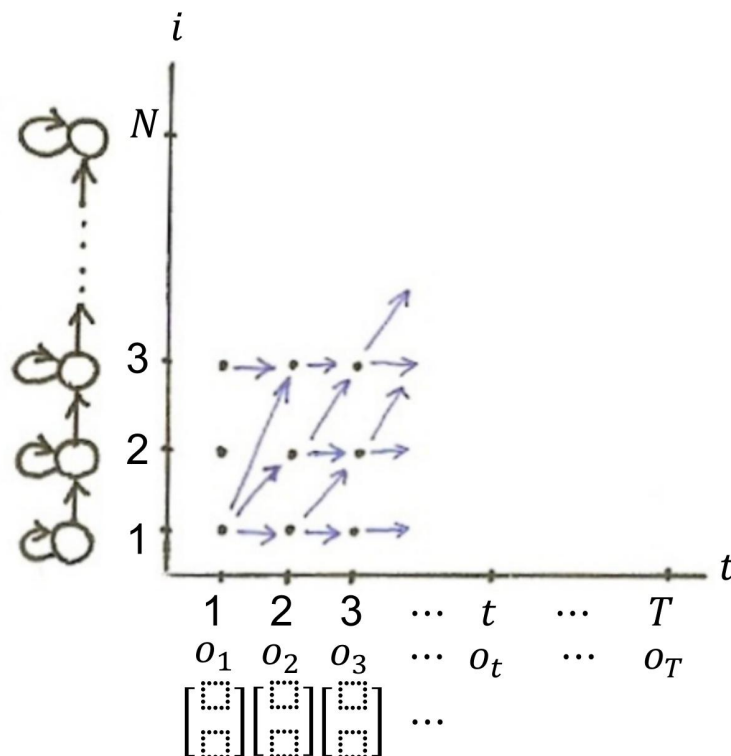
$$[\pi_{q_1} \cdot a_{q_1 q_2} \cdot a_{q_2 q_3} \cdot \dots \cdot a_{q_{T-1} q_T}])$$



$$P(\bar{q}|\lambda)$$

total number of different $\bar{q} : N^T$

huge computation requirements



纵坐标代表隐状态，横坐标是时间轴， o_1, o_2, \dots, o_t 指的是第 t 个时间点第观察状态。
将概率问题转化为路径问题。转化为路径之后就想到了二维表上第动态规划，可以优化那个高阶时间复杂度第计算问题。

Basic Problem 1 for HMM

- Forward Algorithm: defining a forward variable $\alpha_t(i)$

$$\begin{aligned}\alpha_t(i) &= P(o_1 o_2 \dots o_t, q_t = i | \lambda) \\ &= \text{Prob}[\text{observing } o_1 o_2 \dots o_t, \text{ state } i \text{ at time } t | \lambda]\end{aligned}$$

- Initialization

$$\alpha_1(i) = \pi_i b_i(o_1), \quad 1 \leq i \leq N$$

- Induction

$$\begin{aligned}\alpha_{t+1}(j) &= \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(o_{t+1}) \\ &\quad 1 \leq j \leq N \\ &\quad 1 \leq t \leq T-1\end{aligned}$$

- Termination

$$P(\bar{O} | \lambda) = \sum_{i=1}^N \alpha_T(i)$$

See Fig. 6.5 of Rabiner and Juang

- All state sequences, regardless of how long previously, merge to the N state at each time instant t

Basic Problem 2 for HMM

- Problem 2: Given λ and $\bar{O} = o_1 o_2 \dots o_T$, find a best state sequence $\bar{q} = q_1 q_2 \dots q_T$

- Backward Algorithm : defining a backward variable $\beta_t(i)$

$$\begin{aligned}\beta_t(i) &= P(o_{t+1}, o_{t+2}, \dots, o_T | q_t = i, \lambda) \\ &= \text{Prob}[\text{observing } o_{t+1}, o_{t+2}, \dots, o_T | \text{state } i \text{ at time } t, \lambda]\end{aligned}$$

- Initialization

$$\beta_T(i) = 1, \quad 1 \leq i \leq N \quad \left(\beta_{T-1}(i) = \sum_{j=1}^N a_{ij} b_j(o_T) \right)$$

- Induction

$$\begin{aligned}\beta_t(i) &= \sum_{j=1}^N a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) \\ &\quad t = T-1, T-2, \dots, 2, 1, \quad 1 \leq i \leq N\end{aligned}$$

See Fig. 6.6 of Rabiner and Juang

- Combining Forward/Backward Variables

$$\begin{aligned}P(\bar{O}, q_t = i | \lambda) &= \text{Prob} [\text{observing } o_1, o_2, \dots, o_t, \dots, o_T, q_t = i | \lambda] \\ &= \alpha_t(i) \beta_t(i) \\ P(\bar{O} | \lambda) &= \sum_{i=1}^N P(\bar{O}, q_t = i | \lambda) = \sum_{i=1}^N [\alpha_t(i) \beta_t(i)]\end{aligned}$$

详细解释:

动态规划就是:

- 1) 知道第 t 项和 $t-1$ 项或者更多项之间的递推关系
- 2) 知道初始值

而前向后项都满足这两个条件, 因此可以用动态规划求解。

理解 **hmm** 的第一个问题, 只需要知道是利用了前向、后向算法, 其中前向后向算法是利用动态规划来优化指数的大 O 阶问题即可, 具体公式不需要记住, 因为记住几天就会忘。

同时做了假设, 就如下图所示: **A** 和 **B** 相互独立假设, 意味着观测状态 t 之前和之后是相互独立的。这个假设是不合理的。但是这并不影响 **hmm** 是一个效果好的算法

$$\begin{aligned}
 &P(\bar{O}, q_t = i | \lambda) \\
 &= \text{Prob} [\text{observing } o_1, o_2, \dots, o_t, \dots, o_T, q_t = i | \lambda] \\
 &= \alpha_t(i) \beta_t(i)
 \end{aligned}$$

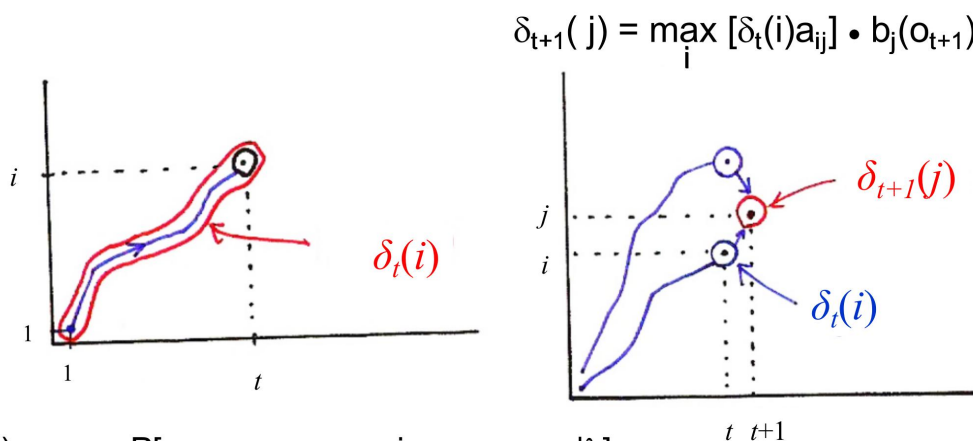
$$\begin{aligned}
 A: &(o_1 o_2 \cdots o_t | \lambda) \\
 B: &(o_{t+1}, o_{t+2}, \cdots o_T | \lambda) \\
 C: &(q_t = i | \lambda) \\
 P(A, B, C) &= P(A, C) P(B|A, C) \\
 // \quad // \quad // \quad (B \perp A) \\
 P(\bar{O}, q_t = i | \lambda) &= \alpha_t(i) P(B|C) \\
 &= \alpha_t(i) \beta_t(i)
 \end{aligned}$$

第二个问题 解码问题

维特比算法。维特比算法其实就是动态规划算法。动态规划就是找第 $t+1$ 项跟前面几项或者前 1 项的关系, 在 **hmm** 里面是跟前面第 t 项之间的关系。解码问题其实就是求当前观测序列条件下最可能的隐状态, 用概率表达, 其实就是最大联合概率密度下式中的 **delta**。

接下来就是求 **delta** 第 $t+1$ 项和第 t 项之间的一种关系 (动态规划的套路), 然后求解初始值 (动态规划的套路), 然后就可以迭代或者递归求第 t 项的值。

1. 规律。转化为路径规划问题之后很容易发现规律
2. 初始值。因为参数都是已知的, 所以 $\text{delta}_0(i)$ 就是 pi , **A** 和 **B** 也都是已知的, 所以可以用维特比来解码了



$$\delta_t(i) = \max_{q_1, q_2, \dots, q_{t-1}} P[q_1, q_2, \dots, q_{t-1}, q_t = i, o_1, o_2, \dots, o_t | \lambda]$$

第三个问题：参数求解问题

参数求解用到了动态规划（前向后向）+em 算法。

和 hmm 的第一个问题很像，第三个问题的直接想法就是通过调整参数来最大化观测序列

Basic Problem 3 for HMM

- Problem 3: Give \bar{O} and an initial model $\lambda = (A, B, \pi)$, adjust λ to maximize $P(\bar{O} | \lambda)$

- Baum-Welch Algorithm (Forward-backward Algorithm)

- Define a new variable

$$\begin{aligned} \xi_t(i, j) &= P(q_t = i, q_{t+1} = j | \bar{O}, \lambda) \\ &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{i=1}^N \sum_{j=1}^N [\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)]} \\ &= \frac{\text{Prob}[\bar{O}, q_t = i, q_{t+1} = j | \lambda]}{P(\bar{O} | \lambda)} \end{aligned}$$

See Fig. 6.7 of Rabiner and Juang

- Recall $\gamma_t(i) = P(q_t = i | \bar{O}, \lambda)$

$\sum_{t=1}^{T-1} \gamma_t(i)$ = expected number of times that state i is visited in \bar{O} from $t = 1$ to $t = T-1$
 = expected number of transitions from state i in \bar{O}

$\sum_{t=1}^{T-1} \xi_t(i, j)$ = expected number of transitions from state i to state j in \bar{O}

em 算法听起来很吓人，其实不过就是两部分，期望部分和最大化部分。

通过引入下面两个变量来看这两个问题。为啥引入这两个参数？因为目标参数都可以用这两个东西表示。同时这两个东西的计算又用到了所要求的参数，我中有你，你中有我，就满足了 em 算法。期望部分是求下面两个变量，最大化部分是更新参数的过程。

因为 hmm 会有一个初始化参数，通过初始化的参数就可以计算出如下两个变量。

$$\gamma_t(i) = \frac{\alpha_t(i) \beta_t(i)}{\sum_{i=1}^N [\alpha_t(i) \beta_t(i)]} = \frac{P(\bar{O}, q_t = i | \lambda)}{P(\bar{O} | \lambda)} = P(q_t = i | \bar{O}, \lambda)$$

$$\begin{aligned} \varepsilon_t(i, j) &= \frac{\alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)}{\sum_{j=1}^N \sum_{i=1}^N \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)} \\ &= \frac{P(\bar{O}, q_t = i, q_{t+1} = j | \lambda)}{P(\bar{O} | \lambda)} = P(q_t = i, q_{t+1} = j | \bar{O}, \lambda) \end{aligned}$$

参数的更新可以通过上面两个变量来计算。

$$\begin{aligned} \bar{\pi}_i &= \gamma_1(i) \\ \bar{a}_{ij} &= \frac{\sum_{t=1}^{T-1} \varepsilon_t(i, j)}{\sum_{t=1}^{T-1} \gamma_t(i)} \\ \bar{b}_j(k) &= \text{Prob}[o_t = v_k \mid q_t = j] = \frac{\sum_{\substack{t=1 \\ o_t = v_k}}^T \gamma_t(j)}{\sum_{t=1}^T \gamma_t(j)} \end{aligned}$$

(for discrete HMM)