隐马尔可夫模型 HMM

贺成

2020/1/14

背景

安德烈·马尔可夫(Андрей Андреевич Марков)

俄国数学家，1856.6-1922.7

主要成就：随机过程（马尔可夫链，马尔可夫过程）

师从

切比雪夫

马尔可夫链

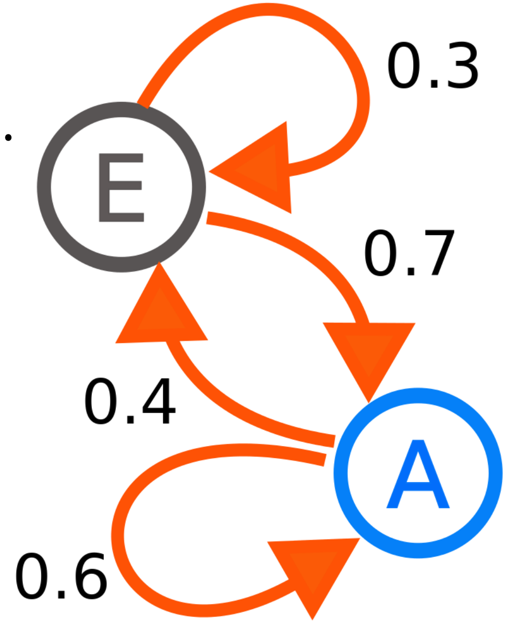
假设有一句话$S = (w_1, w_2, w_3, \dots, w_n)$由n个单词构成，求这句话出现的概率$P(S)$.

$$P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(w_1) \cdot P(w_2|w_1) \cdot P(w_3|w_1, w_2) \cdots P(w_n|w_1, w_2, \cdots w_{n-1})$$

根本没法算!

假设1：每个单词都是独立的，$P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(w_1) \cdot P(w_2) \cdot P(w_3) \cdots P(w_n)$

假设2：每个单词只与前一个单词有关，$P(S) = P(w_1, w_2, w_3, \dots, w_n) = P(w_1) \cdot P(w_2|w_1) \cdot P(w_3|w_2) \cdots P(w_n|w_{n-1})$



假设1 —> 朴素贝叶斯

假设2 —> 马尔可夫假设

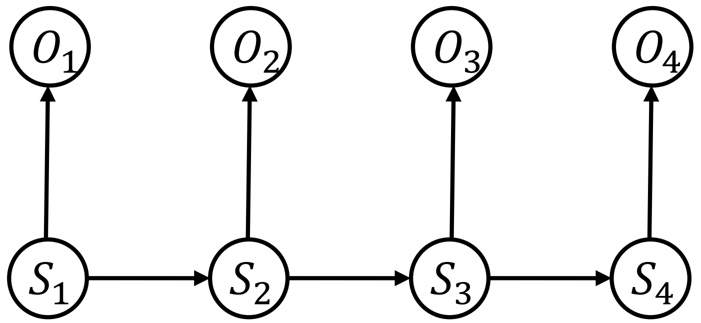
我们把符合马尔可夫假设的随机过程称为马尔可夫链（马尔可夫过程）

隐马尔可夫模型

马尔可夫模型是马尔可夫链的一个扩展，在任意时刻t状态S_t是不可见的，隐藏的。但在每个时刻t会输出一个值O_t，O_t由S_t决定，且只由S_t决定。隐藏的状态$S_1, S_2, \dots, S_t$是一个马尔可夫链。

独立输出假设

马尔可夫假设



马尔可夫模型是一个生成式模型，因此对联合概率建模：

$$P(O,S) = \prod_{T} P(o_t|s_t) \cdot P(s_t|s_{t-1})$$

$P(o_1, o_2, o_3, \cdots |s_1, s_2, s_3, \cdots) = \prod_t P(o_t|s_t)$ (独立输出假设)

$P(s_1, s_2, s_3, \cdots) = \prod_t P(s_t|s_{t-1})$ （马尔可夫假设）

隐马尔可夫模型应用

Baker夫妇

语音识别

20世纪70年代

错误率从30%降到10%

李开复

语音识别

20世纪80年代

世界上第一个大词汇量连续
语音识别系统Sphinx

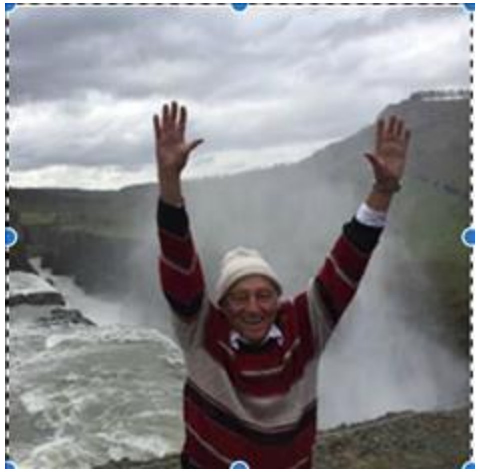
Gary A. Churchill

生物序列

1989年

从那之后，在生物信息学领域HMM逐渐
成为一项不可或缺的技术

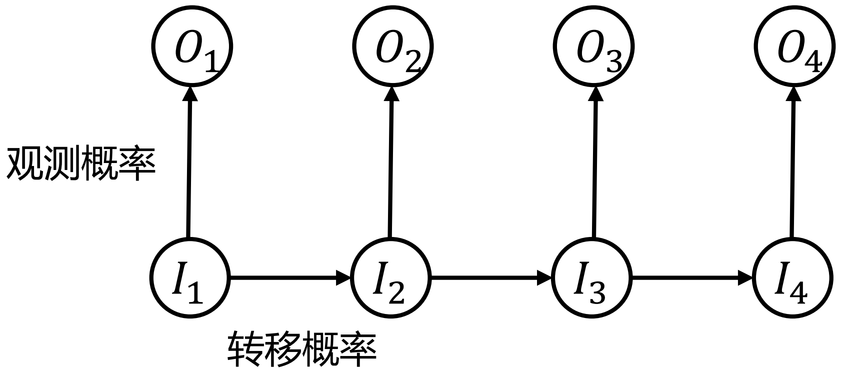
However,…… 隐马尔可夫模型并不是马尔可夫发明的



Leonard E. Baum

而是鲍姆等人在20世纪六七十年代提出的，HMM训练方法中的鲍姆-韦尔奇算法即是以他的名字命名的。

隐马尔可夫模型



- 组成
  - 初始概率分布
  - 状态转移概率分布
  - 观测概率分布
  - Q：所有可能状态的集合
  - V：所有可能观测的集合

$$Q=\{q_1,q_2,\cdots,q_N\}, \quad V=\{v_1,v_2,\cdots,v_M\}$$

  - I: 长度为T的状态序列
  - O：对应的观测序列

$$I=(i_1,i_2,\cdots,i_T), \quad O=(o_1,o_2,\cdots,o_T)$$

隐马尔可夫模型

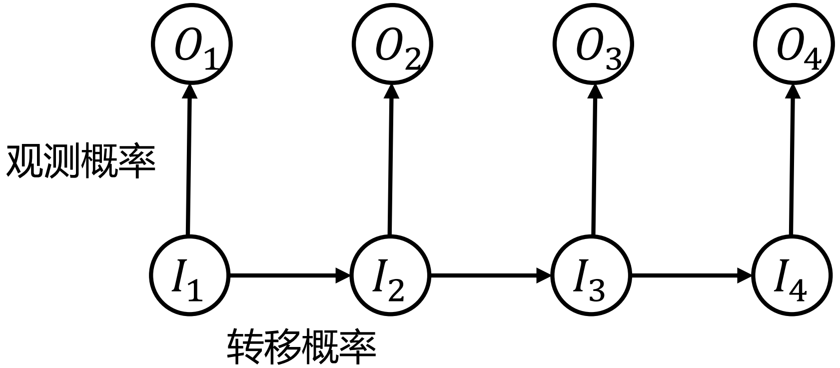
状态转移矩阵A：

$$A = [a_{ij}]_{N \times N}$$

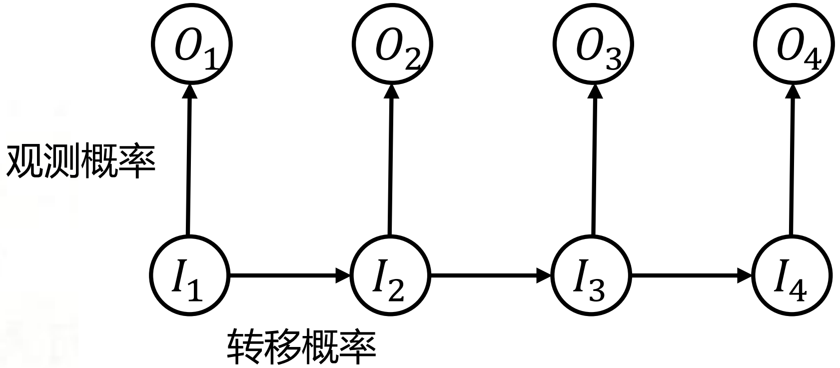
其中，

$$a_{ij} = P(i_{t+1} = q_j | i_t = q_i), \quad i = 1, 2, \cdots, N; \quad j = 1, 2, \cdots, N$$

是在时刻 t 处于状态 q_i 的条件下在时刻 $t+1$ 转移到状态 q_j 的概率。



隐马尔可夫模型



B 是观测概率矩阵：

$$B = [b_j(k)]_{N\times M}$$

其中，

$$b_j(k) = P(o_t = v_k | i_t = q_j), \quad k = 1,2,\cdots,M; \quad j = 1,2,\cdots,N$$

是在时刻 t 处于状态 q_j 的条件下生成观测 v_k 的概率。

π 是初始状态概率向量：

$$\pi = (\pi_i)$$

其中，

$$\pi_i = P(i_1 = q_i), \quad i = 1,2,\cdots,N$$

是时刻 $t=1$ 处于状态 q_i 的概率。

隐马尔可夫模型

隐马尔可夫模型由初始状态概率向量 π、状态转移概率矩阵 A 和观测概率矩阵 B 决定。π 和 A 决定状态序列，B 决定观测序列。因此，隐马尔可夫模型 λ 可以用三元符号表示，即

$$\lambda = (A, B, \pi) \tag{10.7}$$

A, B, π 称为隐马尔可夫模型的三要素。

两个基本假设：

- 隐藏状态符合马尔可夫假设，即任意时刻的状态只依赖于前一状态，和其他状态以及观测状态都无关
- 观测独立性假设，即任意时刻观测只依赖于该时刻的隐状态，与其他观测状态无关。

例子

例 10.1（盒子和球模型） 假设有 4 个盒子，每个盒子里都装有红、白两种颜色的球，盒子里的红、白球数由表 10.1 列出。

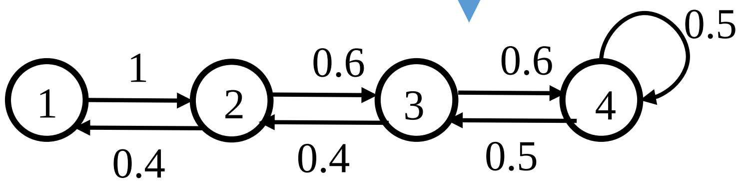
表 10.1　各盒子的红、白球数

	盒子			
	1	2	3	4
红球数	5	3	6	8
白球数	5	7	4	2

按照下面的方法抽球，产生一个球的颜色的观测序列：

- 开始，从 4 个盒子里以等概率随机选取 1 个盒子，从这个盒子里随机抽出 1 个球，记录其颜色后，放回；
- 然后，从当前盒子随机转移到下一个盒子，规则是：如果当前盒子是盒子 1，那么下一盒子一定是盒子 2；如果当前是盒子 2 或 3，那么分别以概率 0.4 和 0.6 转移到左边或右边的盒子；如果当前是盒子 4，那么各以 0.5 的概率停留在盒子 4 或转移到盒子 3；
- 确定转移的盒子后，再从这个盒子里随机抽出 1 个球，记录其颜色，放回；
- 如此下去，重复进行 5 次，得到一个球的颜色的观测序列：

$$O=(\text{红},\ \text{红},\ \text{白},\ \text{白},\ \text{红})$$



Task：根据前面隐马尔可夫模型的定义写出状态集合Q，观测集合V，初始概率分布π，状态转移概率分布A，观测概率分布B。

盒子对应状态，状态的集合是：

$$Q=\{\text{盒子 1},\ \text{盒子 2},\ \text{盒子 3},\ \text{盒子 4}\},\quad N=4$$

球的颜色对应观测。观测的集合是：

$$V=\{\text{红},\ \text{白}\},\quad M=2$$

状态转移概率分布为

$$A=\begin{array}{c} \\ \text{盒1} \\ \text{盒2} \\ \text{盒3} \\ \text{盒4} \end{array}\begin{array}{c} \begin{array}{cccc} \text{盒1} & \text{盒2} & \text{盒3} & \text{盒4} \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix} \end{array}$$

观测概率分布为

$$B=\begin{array}{c} \\ \text{盒1} \\ \text{盒2} \\ \text{盒3} \\ \text{盒4} \end{array}\begin{array}{c} \begin{array}{cc} \text{红球} & \text{白球} \end{array} \\ \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix} \end{array}$$

HMM观测序列的生成过程

算法 10.1（观测序列的生成）

输入：隐马尔可夫模型 $\lambda=(A,B,\pi)$，观测序列长度 T；

输出：观测序列 $O=(o_1,o_2,\cdots,o_T)$。

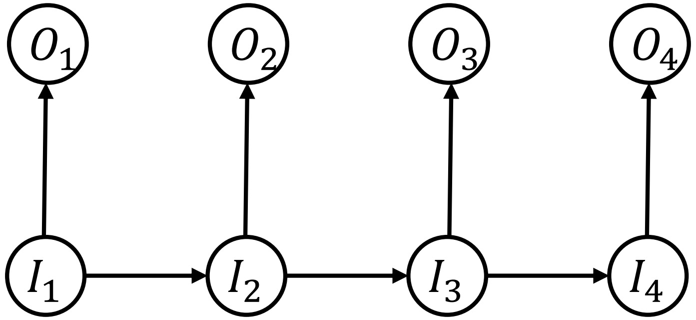
（1）按照初始状态分布 π 产生状态 i_1；

（2）令 $t=1$；

（3）按照状态 i_t 的观测概率分布 $b_{i_t}(k)$ 生成 o_t；

（4）按照状态 i_t 的状态转移概率分布 $\{a_{i_t i_{t+1}}\}$ 产生状态 i_{t+1}，$i_{t+1}=1,2,\cdots,N$；

（5）令 $t=t+1$；如果 $t<T$，转步 (3)；否则，终止。 ■



隐马尔可夫模型的三个基本问题

1、概率计算问题：

给定一个模型如何计算某个特定的输出序列的概率，即计算$P(O|\lambda)$

2、学习问题：

给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

3、预测问题（解码）：

给定一个模型和某个特定的输出序列，如何找到最可能产生这个输出的隐状态序列，即求使得$P(I|O)$最大的状态序列 $I = (i_1, i_2, \cdots i_T)$

概率计算问题：给定一个模型，计算某个特定观测序列O的概率$P(O|\lambda)$

1. 直接计算法

概率公式直接算：$P(O|\lambda)=\sum_I P(O,I|\lambda)=\sum_I P(O|I,\lambda)P(I|\lambda)$

计算$P(O|I,\lambda)$:

对固定的状态序列 $I=(i_1,i_2,\cdots,i_T)$，观测序列 $O=(o_1,o_2,\cdots,o_T)$ 的概率是：

$$P(O|I,\lambda)=b_{i_1}(o_1)b_{i_2}(o_2)\cdots b_{i_T}(o_T)$$

计算$P(I|\lambda)$:　状态序列 $I=(i_1,i_2,\cdots,i_T)$ 的概率是：

$$P(I|\lambda)=\pi_{i_1}a_{i_1i_2}a_{i_2i_3}\cdots a_{i_{T-1}i_T}$$

最终得到$P(O|\lambda)$:

然后，对所有可能的状态序列 I 求和，得到观测序列 O 的概率 $P(O|\lambda)$，即

$$\begin{aligned}P(O|\lambda)&=\sum_I P(O|I,\lambda)P(I|\lambda)\\&=\sum_{i_1,i_2,\cdots,i_T}\pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\cdots a_{i_{T-1}i_T}b_{i_T}(o_T)\end{aligned}$$

时间复杂度 $O(TN^T)$

观测序列长度　　状态个数

概率计算问题：计算观测序列O出现的概率$P(O|\lambda)$

2. 前向算法

首先定义前向概率。

定义 10.2（前向概率） 给定隐马尔可夫模型 λ，定义到时刻 t 部分观测序列为 $o_1, o_2, \cdots, o_t$ 且状态为 q_i 的概率为前向概率，记作

$$\alpha_t(i) = P(o_1, o_2, \cdots, o_t, i_t = q_i|\lambda) \tag{10.14}$$

可以递推地求得前向概率 $\alpha_t(i)$ 及观测序列概率 $P(O|\lambda)$。

算法 10.2（观测序列概率的前向算法）

输入：隐马尔可夫模型 λ，观测序列 O；

输出：观测序列概率 $P(O|\lambda)$。

（1）初值

$$\alpha_1(i) = \pi_i b_i(o_1), \quad i = 1, 2, \cdots, N \tag{10.15}$$

（2）递推 对 $t = 1, 2, \cdots, T-1$，

$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i(o_{t+1}), \quad i = 1, 2, \cdots, N \tag{10.16}$$

（3）终止

$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i) \tag{10.17}$$

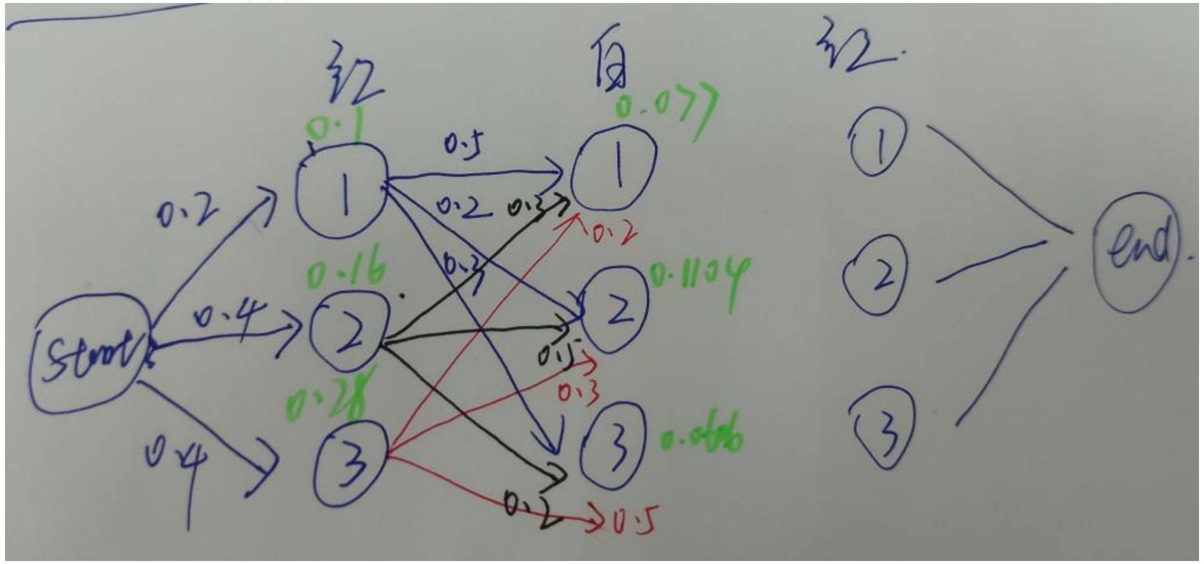
概率计算问题：计算观测序列O出现的概率$P(O|\lambda)$

2. 前向算法——例子

例 10.2 考虑盒子和球模型 $\lambda=(A,B,\pi)$，状态集合 $Q=\{1,2,3\}$，观测集合 $V=\{红, 白\}$，

$$A=\begin{array}{c} \\ 盒1 \\ 盒2 \\ 盒3 \end{array}\begin{array}{c} \begin{array}{ccc} 盒1 & 盒2 & 盒3 \end{array} \\ \left[\begin{array}{ccc} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{array}\right] \end{array}, \quad B=\begin{array}{c} \\ 盒1 \\ 盒2 \\ 盒3 \end{array}\begin{array}{c} \begin{array}{cc} 红 & 白 \end{array} \\ \left[\begin{array}{cc} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{array}\right] \end{array}, \quad \pi=\left[\begin{array}{c} 0.2 \\ 0.4 \\ 0.4 \end{array}\right]$$

设 $T=3$，$O=(红, 白, 红)$，试用前向算法计算 $P(O|\lambda)$。



时间复杂度 $O(TN^2)$

概率计算问题：计算观测序列O出现的概率$P(O|\lambda)$

3. 后向算法

定义 10.3（后向概率） 给定隐马尔可夫模型 λ，定义在时刻 t 状态为 q_i 的条件下，从 $t+1$ 到 T 的部分观测序列为 $o_{t+1}, o_{t+2}, \cdots, o_T$ 的概率为后向概率，记作

$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T | i_t = q_i, \lambda) \tag{10.18}$$

可以用递推的方法求得后向概率 $\beta_t(i)$ 及观测序列概率 $P(O|\lambda)$。

算法 10.3（观测序列概率的后向算法）

输入：隐马尔可夫模型 λ，观测序列 O；

输出：观测序列概率 $P(O|\lambda)$。

（1）

$$\beta_T(i) = 1, \quad i = 1, 2, \cdots, N \tag{10.19}$$

（2）对 $t = T-1, T-2, \cdots, 1$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad i = 1, 2, \cdots, N \tag{10.20}$$

（3）

$$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i) \tag{10.21}$$

■

概率计算问题：计算观测序列O出现的概率$P(O|\lambda)$

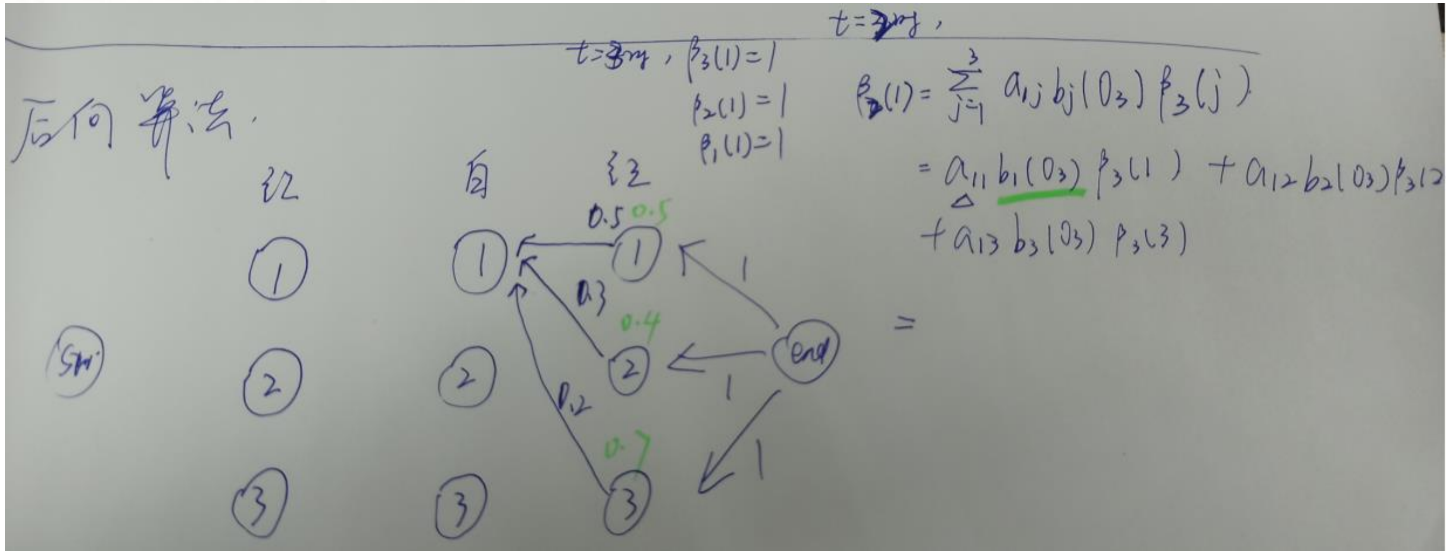
2. 后向算法——例子

例 10.2 考虑盒子和球模型 $\lambda=(A,B,\pi)$，状态集合 $Q=\{1,2,3\}$，观测集合 $V=\{$红, 白$\}$，

$$A=\begin{array}{c}\\ \text{盒1}\\ \text{盒2}\\ \text{盒3}\end{array}\begin{array}{c}\begin{array}{ccc}\text{盒1} & \text{盒2} & \text{盒3}\end{array}\\ \left[\begin{array}{ccc}0.5 & 0.2 & 0.3\\ 0.3 & 0.5 & 0.2\\ 0.2 & 0.3 & 0.5\end{array}\right]\end{array},\quad B=\begin{array}{c}\\ \text{盒1}\\ \text{盒2}\\ \text{盒3}\end{array}\begin{array}{c}\begin{array}{cc}\text{红} & \text{白}\end{array}\\ \left[\begin{array}{cc}0.5 & 0.5\\ 0.4 & 0.6\\ 0.7 & 0.3\end{array}\right]\end{array},\quad \pi=\left[\begin{array}{c}0.2\\ 0.4\\ 0.4\end{array}\right]$$

设 $T=3$，$O=($红, 白, 红$)$，试用前向算法计算 $P(O|\lambda)$。

时间复杂度 $O(TN^2)$



HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、有监督方式训练

假设已给训练数据包含 S 个长度相同的观测序列和对应的状态序列 $\{(O_1, I_1), (O_2, I_2), \cdots, (O_S, I_S)\}$

根据条件概率，我们知道：

$$P(O_t|I_t) = \frac{P(O_t, I_t)}{P(I_t)}$$

$$P(I_t|I_{t-1}) = \frac{P(I_t, I_{t-1})}{P(I_{t-1})}$$

直接去统计:

$$P(O_t|I_t) = \frac{\#(O_t, I_t)}{\#(I_t)}$$

$$P(I_t|I_{t-1}) = \frac{\#(I_t, I_{t-1})}{\#(I_{t-1})}$$

Drawbacks:

需要大量的人工标注，代价高

HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

Baum-Welch算法是一种EM算法

E步，求Q函数：

$$Q(\lambda, \bar{\lambda}) = E_Z[\log(P(O, I|\lambda))|O, \bar{\lambda}]$$

$$= \sum_I \log P(O, I|\lambda) \cdot \boxed{P(I|O, \bar{\lambda})}$$

$$P(I|O, \bar{\lambda}) = \frac{P(I, O|\bar{\lambda})}{\boxed{P(O|\bar{\lambda})}}$$

常数

$$P(O, I|\lambda) = \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{T-1} i_T} b_{i_T}(o^T)$$

$$= \pi_{i_1} \cdot \prod_{t=1}^{T-1} a_{i_t i_{t+1}} \cdot \prod_{t=1}^{T} b_{i_t}(o^t)$$

$$\Rightarrow Q(\lambda, \bar{\lambda}) = \sum_I \log\left(\pi_{i_1} \cdot \prod_{t=1}^{T-1} a_{i_t i_{t+1}} \cdot \prod_{t=1}^{T} b_{i_t}(o^t)\right) \cdot P(O, I|\bar{\lambda})$$

于是函数 $Q(\lambda, \bar{\lambda})$ 可以写成：

$$Q(\lambda, \bar{\lambda}) = \sum_I \log \pi_{i_1} P(O, I|\bar{\lambda}) + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}}\right) P(O, I|\bar{\lambda}) + \sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t)\right) P(O, I|\bar{\lambda}) \tag{10.34}$$

$$Q(\lambda, \bar{\lambda}) = \sum_I \log P(O, I|\lambda) P(O, I|\bar{\lambda})$$

其中，$\bar{\lambda}$ 是隐马尔可夫模型参数的当前估计值，λ 是要极大化的隐马尔可夫模型参数。

$$P(O, I|\lambda) = \pi_{i_1} b_{i_1}(o_1) a_{i_1 i_2} b_{i_2}(o_2) \cdots a_{i_{T-1} i_T} b_{i_T}(o_T)$$

HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

M步：

$$Q(\lambda, \bar{\lambda}) = \boxed{\sum_I \log \pi_{i_1} P(O, I|\bar{\lambda})} + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}}\right) P(O, I|\bar{\lambda}) +$$

$$\sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t)\right) P(O, I|\bar{\lambda}) \tag{10.34}$$

$$\sum_I \log \pi_{i_1} P(O, I|\bar{\lambda}) = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_T} [\log \pi_{i_1} P(O, i_1, i_2, \cdots i_T|\bar{\lambda})]$$

$$= \sum_{i_1} \log \pi_{i_1} P(O, i_1|\bar{\lambda})$$

$$= \sum_{j=1}^{N} \log \pi_j P(O, i_1 = j|\bar{\lambda})$$

$$\sum_I \log \pi_{i_1} P(O, I|\bar{\lambda}) = \sum_{i=1}^{N} \log \pi_i P(O, i_1 = i|\bar{\lambda})$$

注意到 π_i 满足约束条件 $\sum_{i=1}^{N} \pi_i = 1$，利用拉格朗日乘子法，写出拉格朗日函数：

$$\sum_{i=1}^{N} \log \pi_i P(O, i_1 = i|\bar{\lambda}) + \gamma \left(\sum_{i=1}^{N} \pi_i - 1\right)$$

对其求偏导数并令结果为 0

$$\frac{\partial}{\partial \pi_i}\left[\sum_{i=1}^{N} \log \pi_i P(O, i_1 = i|\bar{\lambda}) + \gamma \left(\sum_{i=1}^{N} \pi_i - 1\right)\right] = 0 \tag{10.35}$$

得

$$P(O, i_1 = i|\bar{\lambda}) + \gamma \pi_i = 0$$

对 i 求和得到 γ

$$\gamma = -P(O|\bar{\lambda})$$

代入式 (10.35) 即得

$$\pi_i = \frac{P(O, i_1 = i|\bar{\lambda})}{P(O|\bar{\lambda})} \tag{10.36}$$

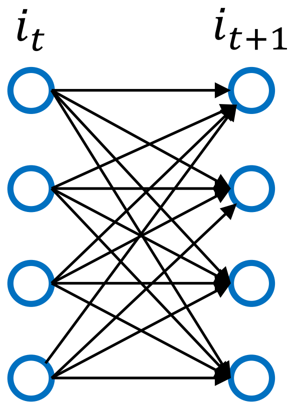
HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

M步：

$$Q(\lambda, \bar{\lambda}) = \sum_I \log \pi_{i_1} P(O, I|\bar{\lambda}) + \boxed{\sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(O, I|\bar{\lambda})} + \sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t) \right) P(O, I|\bar{\lambda}) \qquad (10.34)$$

$$\sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}} \right) P(O, I|\bar{\lambda}) = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \log a_{ij} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})$$



类似第 1 项，应用具有约束条件 $\sum_{j=1}^{N} a_{ij} = 1$ 的拉格朗日乘子法可以求出

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i|\bar{\lambda})} \qquad (10.37)$$

HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

M步：

$$Q(\lambda,\bar{\lambda}) = \sum_I \log \pi_{i_1} P(O,I|\bar{\lambda}) + \sum_I \left(\sum_{t=1}^{T-1} \log a_{i_t i_{t+1}}\right) P(O,I|\bar{\lambda}) + \boxed{\sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t)\right) P(O,I|\bar{\lambda})} \tag{10.34}$$

$$\sum_I \left(\sum_{t=1}^{T} \log b_{i_t}(o_t)\right) P(O,I|\bar{\lambda}) = \sum_{j=1}^{N}\sum_{t=1}^{T} \log b_j(o_t) P(O, i_t = j|\bar{\lambda})$$

同样用拉格朗日乘子法，约束条件是 $\sum_{k=1}^{M} b_j(k) = 1$。注意，只有在 $o_t = v_k$ 时 $b_j(o_t)$ 对 $b_j(k)$ 的偏导数才不为 0，以 $I(o_t = v_k)$ 表示。求得

$$b_j(k) = \frac{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda}) I(o_t = v_k)}{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda})} \tag{10.38}$$

HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda = (A, B, \pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

参数估计公式

令：

$$\gamma_t(i) = P(i_t = q_i | O, \lambda) = \frac{P(i_t = q_i, O|\lambda)}{P(O|\lambda)}$$

$$\xi_t(i,j) = P(i_t = q_i, i_{t+1} = q_j | O, \lambda)$$

$$b_j(k) = \frac{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda}) I(o_t = v_k)}{\sum_{t=1}^{T} P(O, i_t = j|\bar{\lambda})} \longrightarrow b_j(k) = \frac{\sum_{t=1, o_t = v_k}^{T} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

$$a_{ij} = \frac{\sum_{t=1}^{T-1} P(O, i_t = i, i_{t+1} = j|\bar{\lambda})}{\sum_{t=1}^{T-1} P(O, i_t = i|\bar{\lambda})} \longrightarrow a_{ij} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{t=1}^{T-1} \gamma_t(i)}$$

$$\pi_i = \frac{P(O, i_1 = i|\bar{\lambda})}{P(O|\bar{\lambda})} \longrightarrow \pi_i = \gamma_1(i)$$

HMM学习问题：给足够量的观测数据，如何估计出HMM模型的参数$\lambda=(A,B,\pi)$

1、无监督方式训练——Baum-Welch （鲍姆-韦尔奇）

Baum-Welch算法流程

算法 10.4（Baum-Welch 算法）

输入：观测数据 $O=(o_1,o_2,\cdots,o_T)$；

输出：隐马尔可夫模型参数。

（1）初始化。对 $n=0$，选取 $a_{ij}^{(0)}$，$b_j(k)^{(0)}$，$\pi_i^{(0)}$，得到模型 $\lambda^{(0)}=(A^{(0)},B^{(0)},\pi^{(0)})$。

（2）递推。对 $n=1,2,\cdots,$

$$a_{ij}^{(n+1)}=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)}$$

$$b_j(k)^{(n+1)}=\frac{\sum_{t=1,o_t=v_k}^{T}\gamma_t(j)}{\sum_{t=1}^{T}\gamma_t(j)}$$

$$\pi_i^{(n+1)}=\gamma_1(i)$$

右端各值按观测 $O=(o_1,o_2,\cdots,o_T)$ 和模型 $\lambda^{(n)}=(A^{(n)},B^{(n)},\pi^{(n)})$ 计算。式中 $\gamma_t(i)$，$\xi_t(i,j)$ 由式 (10.24) 和式 (10.26) 给出。

（3）终止。得到模型参数 $\lambda^{(n+1)}=(A^{(n+1)},B^{(n+1)},\pi^{(n+1)})$。■

HMM预测问题（解码）

给定训练好的HMM模型和某个特定的输出序列，

如何找到最有可能产生这个输出的隐状态序列，即求使得$P(I|O)$最大的状态序列 $I=(i_1,i_2,\cdots i_T)$?

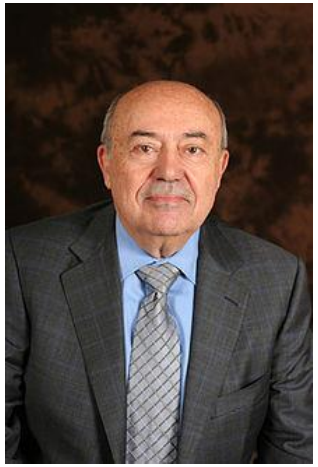
- 暴力枚举，复杂度太高，不可行。
- 贪心算法，在每个时刻t转移到概率最大的下一个状态，不能保证全局最优。
- 动态规划——Viterbi（维特比）算法

HMM预测问题（解码）

给定训练好的HMM模型和某个特定的输出序列，

如何找到最有可能产生这个输出的隐状态序列，即求使得$P(I|O)$最大的状态序列 $I=(i_1,i_2,\cdots i_T)$?

维特比算法



1935年3月9日

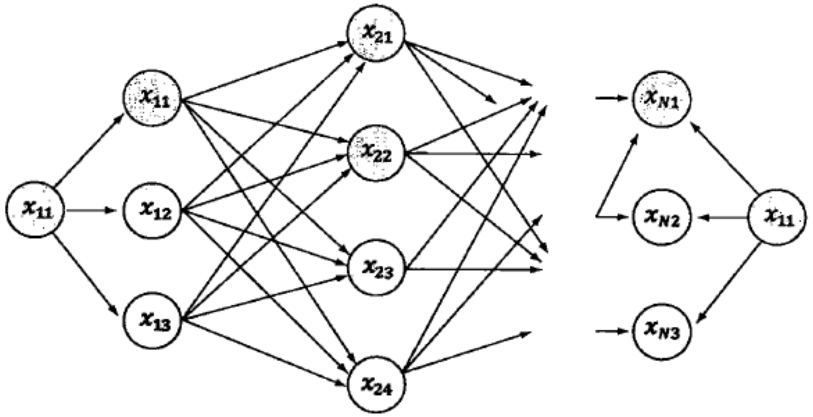
Qualcomm

维特比算法的基础可以概括成下面三点：

1. 如果概率最大的路径P（或者说最短路径）经过某个点，比如图中的x_{22}，那么这条路径上从起始点S到x_{22}的这一段子路径Q，一定是S到x_{22}之间的最短路径。否则，用S到x_{22}的最短路径R替代Q，便构成了一条比P更短的路径，这显然是矛盾的。

2. 从S到E的路径必定经过第i时刻的某个状态（这显然是大白话，但是很关键），假定第i时刻有k个状态，那么如果记录了从S到第i个状态的所有k个节点的最短路径，最终的最短路径必经过其中的一条。这样，在任何时刻，只要考虑非常有限条最短路径即可。

3. 结合上述两点，假定当我们从状态i进入状态$i+1$时，从S到状态i上各个节点的最短路径已经找到，并且记录在这些节点上，那么在计算从起点S到第$i+1$状态的某个节点x_{i+1}的最短路径时，只要考虑从S到前一个状态i所有的k个节点的最短路径，以及从这k个节点到x_{i+1}，j的距离即可。



HMM预测问题（解码）

给定训练好的HMM模型和某个特定的输出序列，

如何找到最有可能产生这个输出的隐状态序列，即求使得$P(I|O)$最大的状态序列 $I=(i_1,i_2,\cdots i_T)$?

算法 10.5（维特比算法）

输入：模型 $\lambda=(A,B,\pi)$ 和观测 $O=(o_1,o_2,\cdots,o_T)$；

输出：最优路径 $I^*=(i_1^*,i_2^*,\cdots,i_T^*)$。

（1）初始化

$$\delta_1(i)=\pi_i b_i(o_1),\quad i=1,2,\cdots,N$$

$$\Psi_1(i)=0,\quad i=1,2,\cdots,N$$

（2）递推。对 $t=2,3,\cdots,T$

$$\delta_t(i)=\max_{1\leqslant j\leqslant N}[\delta_{t-1}(j)a_{ji}]b_i(o_t),\quad i=1,2,\cdots,N$$

$$\Psi_t(i)=\arg\max_{1\leqslant j\leqslant N}[\delta_{t-1}(j)a_{ji}],\quad i=1,2,\cdots,N$$

（3）终止

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

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

（4）最优路径回溯。对 $t=T-1,T-2,\cdots,1$

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

求得最优路径 $I^*=(i_1^*,i_2^*,\cdots,i_T^*)$。

HMM预测问题（解码）

给定训练好的HMM模型和某个特定的输出序列，

如何找到最有可能产生这个输出的隐状态序列，即求使得$P(I|O)$最大的状态序列 $I = (i_1, i_2, \cdots i_T)$?

维特比算法——例子

例 10.3 例 10.2 的模型 $\lambda = (A, B, \pi)$,

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

已知观测序列 $O =$ (红，白，红)，试求最优状态序列，即最优路径 $I^* = (i_1^*, i_2^*, i_3^*)$。