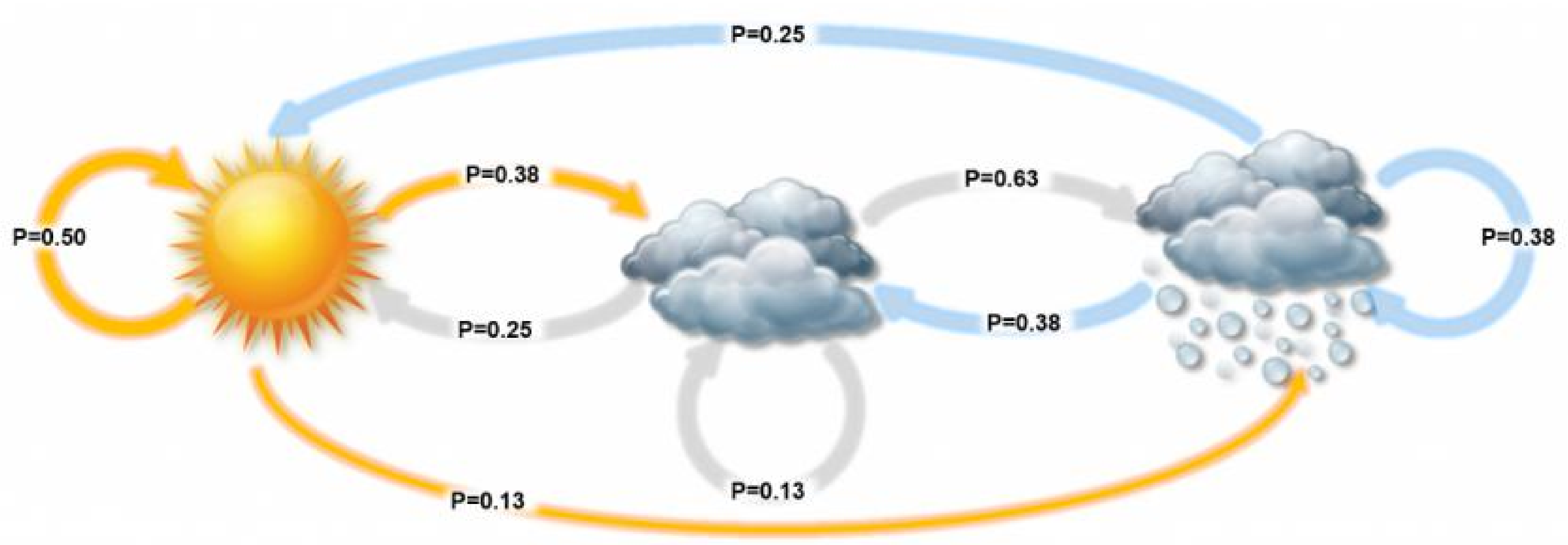
隐马尔科夫HMM

- 马尔科夫模型

  天气变化种类：晴天，多云，雷雨，他们之间应该有些联系吧！



隐马尔科夫HMM

马尔科夫模型

状态之间可以发生转换，昨天和今天转换的情况：

$$
\begin{array}{cc|ccc}
 & & & Today & \\
 & & sun & cloud & rain \\
 & sun & 0.50 & 0.375 & 0.125 \\
Yesterday & cloud & 0.25 & 0.125 & 0.625 \\
 & rain & 0.25 & 0.375 & 0.375
\end{array}
$$

隐马尔科夫HMM

马尔科夫模型

今天能得到明天的情况，明天能得到后天的情况，以此类推可以无限的玩下去那是不是得有一个初始的情况才能一直玩下去啊！

$$\begin{array}{ccc} sun & cloud & rain \\ [1.0 & 0.0 & 0.0] \end{array}$$

这里我们就定义好了一个一阶马尔科夫模型：

状态：晴天，多云，雷雨

状态转换概率：三种天气状态间的转换概率

初始概率：晴天

隐马尔科夫HMM

马尔科夫模型

计算今天(t=1)的天气状况：

今天为晴天的概率=初始晴天概率X晴天转晴天概率

+初始多云概率X多云转晴天概率

+初始雷雨概率X雷雨转晴天概率。

t=1

weather	P
Sun	0.500
Cloud	0.375
Rain	0.125

t=2

weather	P
Sun	0.375
Cloud	0.281
Rain	0.344

t=3

weather	P
Sun	0.344
Cloud	0.305
Rain	0.352

隐马尔科夫HMM

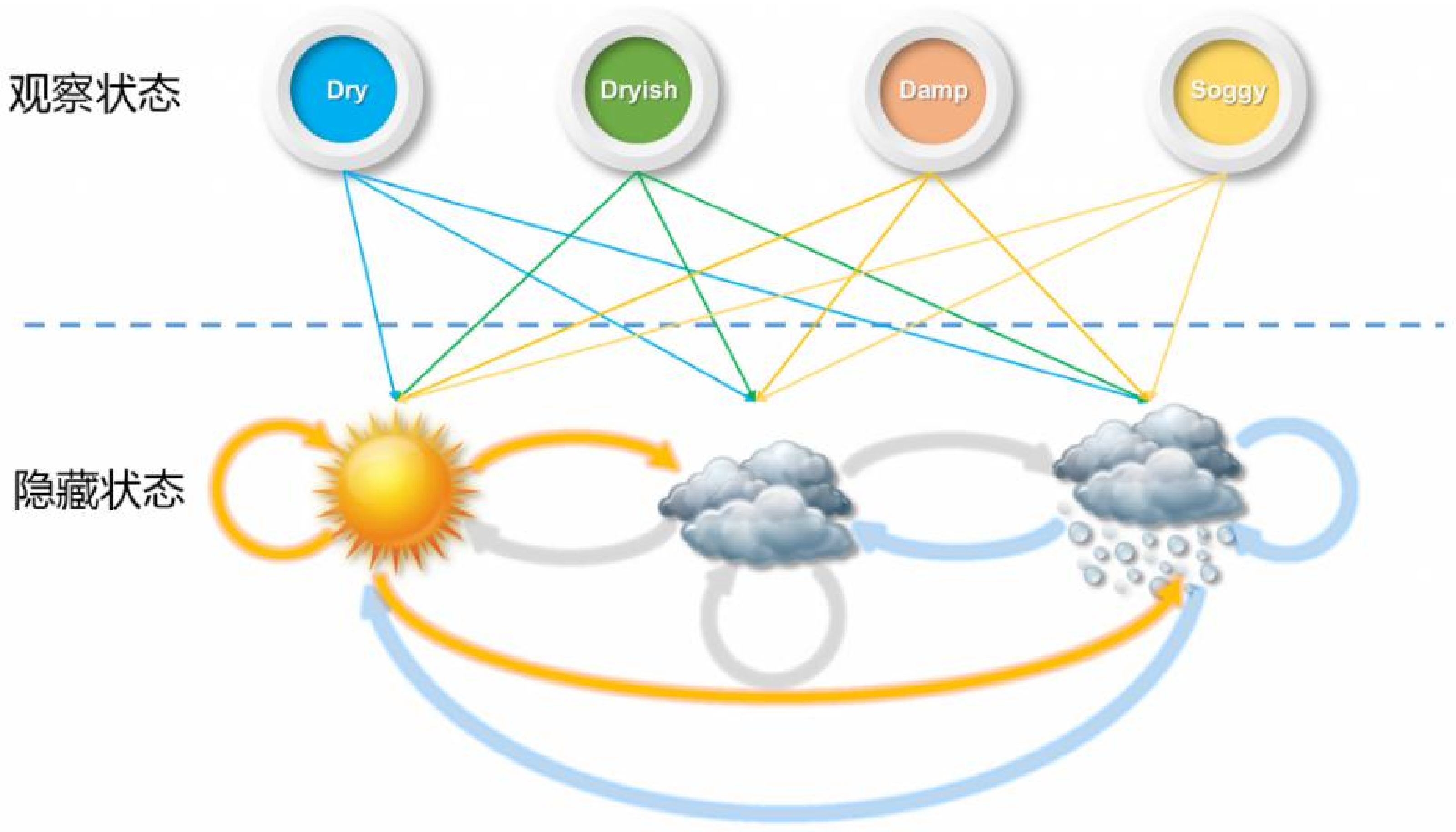
- 游戏难度加大了

- 现在我们漂到了一个岛上，这里没有天气预报，只有一片片的海藻。
- 这些海藻状态能观察到的：

- 这里我们就没有直接的天气信息了，有的是间接的信息，海藻的状态跟天气的变换有一定的关系。
- 既然海藻是能看到的，那它就是**观察状态**，天气信息看不到就是**隐藏状态**。

隐马尔科夫HMM

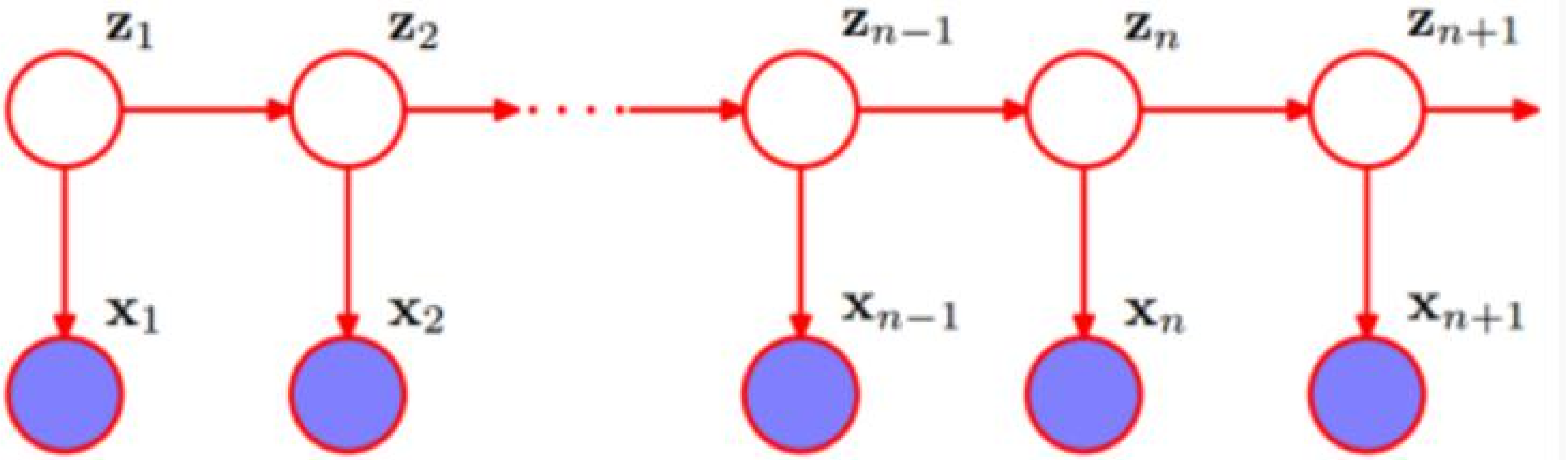
- 隐马尔科夫模型



隐马尔科夫HMM

- 隐马尔科夫模型

  - 当前的状态只和前一状态有关：$P(z_t|z_{t-1},x_{t-1}, z_{t-2}, x_{t-2}, \ldots, z_1, x_1)= P(z_t \mid z_{t-1})$
  - 某个观测只和生成它的状态有关：$P(x_t|z_t,x_t, z_{t-1}, x_{t-1}, z_{t-2}, x_{t-2},\ldots, z_1, x_1) = P(x_t \mid z_t)$



隐马尔科夫HMM

隐马尔科夫模型的组成

- 三个必备：初始概率(π)，隐藏状态转移概率矩阵(A)，生成观测状态概率矩阵(B)。

$$HMM = (\pi, A, B)$$

- 隐藏状态与观察状态（B矩阵）：

$$
\begin{array}{ll}
 & \qquad\qquad Seaweed \\
 & \begin{array}{cccc} Dry & Dryish & Damp & Soggy \end{array} \\
Weather \; \begin{array}{r} sun \\ cloud \\ rain \end{array} & \begin{bmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{bmatrix}
\end{array}
$$

隐马尔科夫HMM

- 要解决的问题:　模型为 $\lambda=(A,B,\pi)$

- 1.给定模型 (π,A,B) 及观测序列 $O=\{o_1,o_2,...o_T\}$ 计算其出现的概率 $P(O|\lambda)$

- 2.给定观测序列 $O=\{o_1,o_2,...o_T\}$，求解参数 (π,A,B) 使得 $P(O|\lambda)$ 最大

- 3.已知模型 (π,A,B) 和观测序列 $O=\{o_1,o_2,...o_T\}$ 求状态序列，使得

  $P(I|O,\lambda)$ 最大

隐马尔科夫HMM

- 求观测序列的概率

- 暴力求解：我们要求的是在给定模型下观测序列出现的概率，那如果我能把所有的隐藏序列都给列出来，也就可以知道联合概率分布 $P(O,I|\lambda)$。

- $P(O|\lambda)=\sum_I P(O,I|\lambda)$，$P(O,I|\lambda)=P(I|\lambda)P(O|I,\lambda)$ 现在要求的目标就很明确了。

- $P(I|\lambda)$ 在给定模型下，一个隐藏序列出现的概率，那就由初始状态慢慢转换嘛。

$I=\{i_1,i_2,\ldots,i_T\}$出现的概率为：$P(I|\lambda)=\pi_{i_1}a_{i_1 i_2}a_{i_2 i_3}\cdots a_{i_{T-1}\ i_T}$

隐马尔科夫HMM

求观测序列的概率

- 对于固定的隐藏序列 $I=\{i_1,i_2,\ldots,i_T\}$，得到观察序列 $O=\{o_1,o_2,\ldots o_T\}$的概率：

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

- 联合概率：$P(O,I|\lambda)=P(I|\lambda)P(O|I,\lambda)=\pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\ldots a_{i_{T-1}\ i_T}b_{i_T}(o_T)$

- 观测序列概率：$P(O|\lambda)=\sum_I P(O,I|\lambda)=\sum_{i_1,i_2,\ldots i_T}\pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\ldots a_{i_{T-1}\ i_T}b_{i_T}(o_T)$

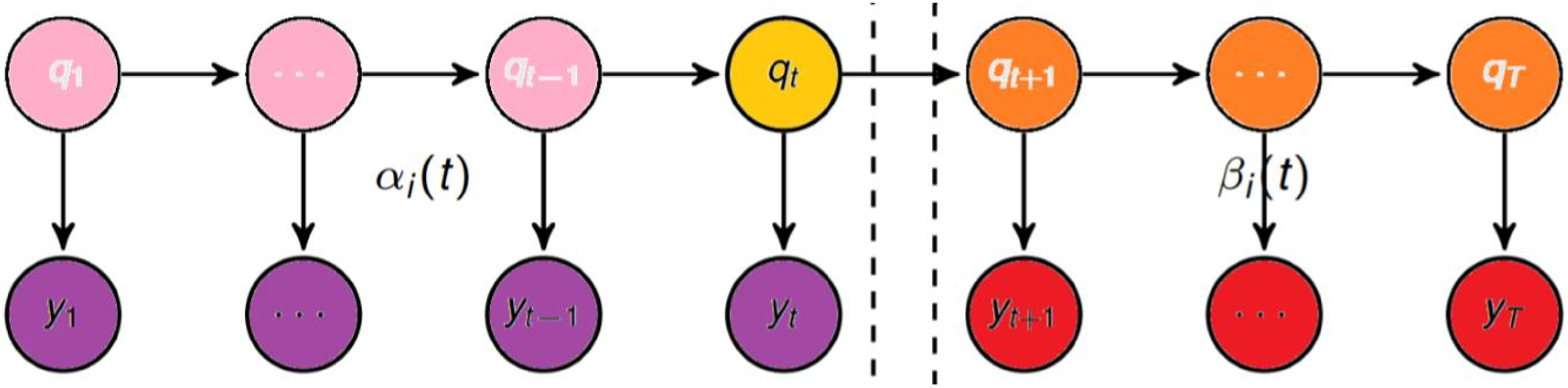
- 复杂度：如果隐藏状态数有N个，$O(TN^T)$

隐马尔科夫HMM

- 前向算法

给定t时刻的隐藏状态为i，观测序列为o1,o2...ot的概率叫做前向概率：

$$\alpha_i(t) = p(y_1, y_2, \ldots y_t, q_t = i|\lambda)$$



隐马尔科夫HMM

前向算法

- 当t=T时，$\alpha_i(T) = p(y_1, y_2, \ldots, y_T, q_T = i \mid \lambda)$ 这表示最后一个时刻，隐藏状态位于第i号状态上并且观测到y1,y2...yT的概率。
- 看起来跟我们要求解的目标有点区别，多了一个第i号状态，如果我能把所有可能的状态都拿过来，不就完事了嘛！$\alpha_1(T) + \alpha_2(T) + \ldots + \alpha_n(T)$
- 那如何求解T时刻的呢？这个看起来是一个动态规划问题了。

隐马尔科夫HMM

前向算法

第一个时刻：$\alpha_i(1) = P(y_1,q_1=i \mid \lambda)$

表示第一时刻的隐藏状态为i，观测序列为y1，$\alpha_i(1) = \pi_i * B_{iy1}$

现在到了第t时刻，状态为j，那t+1时刻状态为i表示为：$a_i(t+1) = (\Sigma\_j\ a_j(t)a_{ji})b_{iy(t+1)}$

其中t时状态为j的前向概率为：$a_j(t)$ 又转移到了状态i：$a_j(t)a_{ji}$

但是这里要考虑t时刻所有的可能：$\Sigma\_j\ a_j(t)\ a_{ji}$ 还要得到t+1的观测。

最终结果：$P(O|\lambda) = \Sigma\_i\ \alpha_i(T)$

隐马尔科夫HMM

前向算法

有3个盒子，每个盒子都有红色和白色两种球，分别为：

盒子1：5红5白

盒子2：4红6白

盒子3：7红3白

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

隐马尔科夫HMM

前向算法

- 得到的观测序列：$O=\{红，白，红\}$
- 观测状态集合：$V=\{红，白\}，M=2$
- 隐藏状态集合：$Q=\{盒子1，盒子2，盒子3\}，N=3$

隐马尔科夫HMM

- 前向算法

时刻1，

(红色球，盒子1) $\alpha_1(1) = \pi_1 b_1(o_1) = 0.2 \times 0.5 = 0.1$

(红色球，盒子2) $\alpha_1(2) = \pi_2 b_2(o_1) = 0.4 \times 0.4 = 0.16$

(红色球，盒子3) $\alpha_1(3) = \pi_3 b_3(o_1) = 0.4 \times 0.7 = 0.28$

时刻2，

(白色球，盒子1)

$$\alpha_2(1) = \left[\sum_{i=1}^{3} \alpha_1(i) a_{i1}\right] b_1(o_2) = [0.1 * 0.5 + 0.16 * 0.3 + 0.28 * 0.2] \times 0.5 = 0.077$$

(白色球，盒子2)

$$\alpha_2(2) = \left[\sum_{i=1}^{3} \alpha_1(i) a_{i2}\right] b_2(o_2) = [0.1 * 0.2 + 0.16 * 0.5 + 0.28 * 0.3] \times 0.6 = 0.1104$$

(白色球，盒子3)

$$\alpha_2(3) = \left[\sum_{i=1}^{3} \alpha_1(i) a_{i3}\right] b_3(o_2) = [0.1 * 0.3 + 0.16 * 0.2 + 0.28 * 0.5] \times 0.3 = 0.0606$$

隐马尔科夫HMM

前向算法

时刻3，（红色球，盒子1） $P(O|\lambda)=\sum_{i=1}^{3}\alpha_3(i)=0.130$

$$\alpha_3(1)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i1}\right]b_1(o_3)=[0.077*0.5+0.1104*0.3+0.0606*0.2]\times 0.5=0.041$$

（红色球，盒子2）

$$\alpha_3(2)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i2}\right]b_2(o_3)=[0.077*0.2+0.1104*0.5+0.0606*0.3]\times 0.4=0.035$$

（红色球，盒子3）

$$\alpha_3(3)=\left[\sum_{i=1}^{3}\alpha_3(i)a_{i3}\right]b_3(o_3)=[0.077*0.3+0.1104*0.2+0.0606*0.5]\times 0.7=0.052$$

隐马尔科夫HMM

- 求解参数：

  - 如果我们有观测序列和其对应的状态序列：$\{(O_1, I_1), (O_2, I_2), \ldots, (O_S, I_S)\}$
    这求解就简单多了，直接计算概率值就可以了！
  - 状态转移概率：$\hat{a}_{ij} = \frac{A_{ij}}{\sum_{j=1}^{N} A_{ij}}, i = 1, 2, \ldots, N, j = 1, 2, \ldots, N$
  - 生成观测概率：$\hat{b}_j(k) = \frac{B_{jk}}{\sum_{k=1}^{M} B_{jk}}, j = 1, 2, \ldots, N; k = 1, 2, \ldots, M$
  - 初始状态概率：查就得了，有数据算起来都容易！

隐马尔科夫HMM

- Baum-Welch算法

  - 给定观测序列$O=\{o_1,o_2,\ldots o_T\}$，求解参数(π,A,B)使得$P(O|\lambda)$最大

  - 这回我们就要估计模型参数，但是由于状态序列未知，相当于是一个含有隐变量的参数估计问题，这就需要EM算法了。

  - Q函数是完全数据的对数似然函数关于给定模型参数和观测变量的前提下对隐变量的条件概率分布的期望，接下来只需要对它进行极大化。

$$Q(\lambda,\overline{\lambda})=\sum_I logP(O,I|\lambda)P(I|O,\overline{\lambda})$$

隐马尔科夫HMM

- Baum-Welch算法

  - 对Q函数进行化简：$Q(\lambda,\overline{\lambda})=\sum_{I} logP(O,I|\lambda)P(I|O,\overline{\lambda})$
  - 其后部分：$P(I|O,\overline{\lambda})=\dfrac{P(O,I|\overline{\lambda})}{P(O|\overline{\lambda})}$ ，在这里对于估计参数来说分母就是常量了。
  - 新的Q函数：$Q(\lambda,\overline{\lambda})=\sum_{I} logP(O,I|\lambda)P(O,I|,\overline{\lambda})$ ，其中

    $$P(O,I|\lambda)=\pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\ldots a_{i_{T-1}i_T}b_{i_T}(o_T)$$

  - 由于这里有对数，这些还是累乘，式子得重新展开了。

隐马尔科夫HMM

Baum-Welch算法

Q函数展开：

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

这么一看，这三个式子里面不就分别是我们想要求解的参数了嘛，既然是加法，那只需要分别最大化就可以了！

以第一项举例：$\sum_{I}\log\pi_{i_1}P(O,I|\hat{\lambda})=\sum_{i=1}^{N}\log\pi_i P(O,i_1=i|\hat{\lambda})$, $\sum_{i=1}^{N}\pi_i=1$

在条件下求极值，拉格朗日乘子法！

隐马尔科夫HMM

Baum-Welch算法

拉格朗日函数： $\sum_{i=1}^{N}\log\pi_i P(O,i_1=i|\hat{\lambda})+\gamma(\sum_{i=1}^{N}\pi_i-1)$

令其偏导为0： $\frac{\partial}{\partial_{\pi_i}}[\sum_{i=1}^{N}\log\pi_i P(O,i_1=i|\hat{\lambda})+\gamma(\sum_{i=1}^{N}\pi_i-1)]=0$

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

对i进行求和： $\gamma=-P(O|\hat{\lambda})$

参数： $\pi_i=\frac{P(O,i_1=i|\hat{\lambda})}{P(O|\hat{\lambda})}$ $a_{ij}=\frac{\sum_{t=1}^{T-1}P(O,i_t=i,i_{t+1}=j|\hat{\lambda})}{\sum_{t=1}^{T-1}P(O,i_t=i|\hat{\lambda})}$ $b_j(k)=\frac{\sum_{t=1}^{T}P(O,i_t=j|\hat{\lambda})I(o_t=v_k)}{\sum_{t=1}^{T}P(O,i_t=j|\hat{\lambda})}$

隐马尔科夫HMM

维特比算法

- 在给定模型 $\lambda=(A,B,\Pi)$ 和观测序列 $O=\{o_1,o_2,\ldots o_T\}$，求最有可能的状态序列 $I^*=\{i_1^*,i_2^*,\ldots i_T^*\}$ 即使得 $P(I^*|O)$ 最大。

- 对于t时刻，隐藏状态为i，要找到所有可能路径的最大值：

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

- 递推公式：$\delta_{t+1}(i)=\max_{i_1,i_2,\ldots i_t} P(i_{t+1}=i,i_1,i_2,\ldots i_t,o_{t+1},o_t,\ldots o_1|\lambda)$

$$=\max_{1\leq j\leq N}[\delta_t(j)a_{ji}]b_i(o_{t+1})$$

- 概率最大路径中t-1的状态：$\Psi_t(i)=arg\max_{1\leq j\leq N}[\delta_{t-1}(j)a_{ji}]$

隐马尔科夫HMM

维特比算法

观测序列：$O=\{红，白，红\}$

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

t1时刻：$\delta_1(1)=\pi_1 b_1(o_1)=0.2\times 0.5=0.1$

$$\delta_1(2)=\pi_2 b_2(o_1)=0.4\times 0.4=0.16$$

$$\delta_1(3)=\pi_3 b_3(o_1)=0.4\times 0.7=0.28$$

$$\Psi_1(1)=\Psi_1(2)=\Psi_1(3)=0$$

隐马尔科夫HMM

- 维特比算法:

t2时刻: $\delta_2(1) = \max_{1 \le j \le 3}[\delta_1(j)a_{j1}]b_1(o_2) = \max_{1 \le j \le 3}[0.1 \times 0.5, 0.16 \times 0.3, 0.28 \times 0.2] \times 0.5 = 0.028$

$$\Psi_2(1) = 3$$

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

$$\Psi_2(2) = 3$$

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

$$\Psi_2(3) = 3$$

隐马尔科夫HMM

维特比算法：

t3时刻：$\delta_3(1) = \max\limits_{1\leq j\leq 3}[\delta_2(j)a_{j1}]b_1(o_3) = \max\limits_{1\leq j\leq 3}[0.028\times 0.5, 0.0504\times 0.3, 0.042\times 0.2]\times 0.5 = 0.00756$

$$\Psi_3(1) = 2$$

$$\delta_3(2) = \max_{1\leq j\leq 3}[\delta_2(j)a_{j2}]b_2(o_3) = \max_{1\leq j\leq 3}[0.028\times 0.2, 0.0504\times 0.5, 0.042\times 0.3]\times 0.4 = 0.01008$$

$$\Psi_3(2) = 2$$

$$\delta_3(3) = \max_{1\leq j\leq 3}[\delta_2(j)a_{j3}]b_3(o_3) = \max_{1\leq j\leq 3}[0.028\times 0.3, 0.0504\times 0.2, 0.042\times 0.5]\times 0.7 = 0.0147$$

$$\Psi_3(3) = 3$$

隐藏序列结果：$(3,3,3)$