隐马尔可夫模型与条件随机场笔记

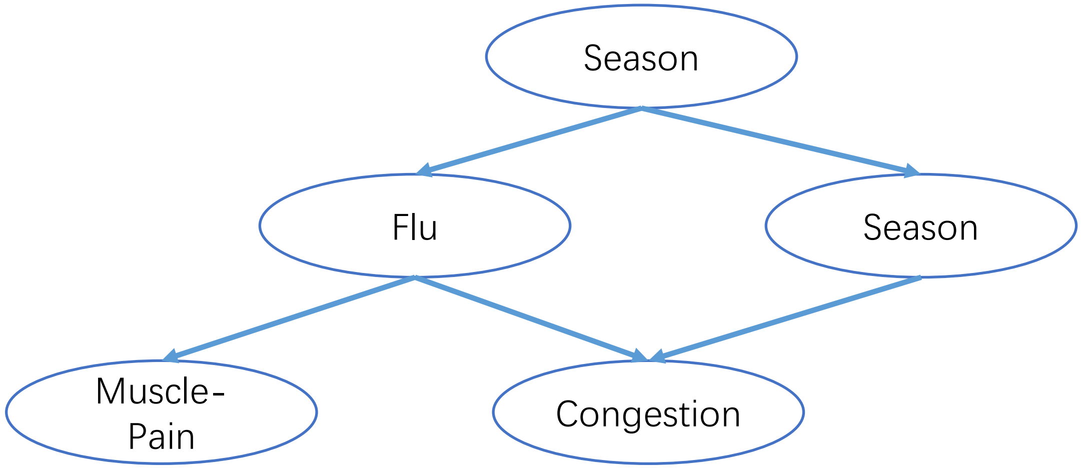
baiziyu

2018年6月30日

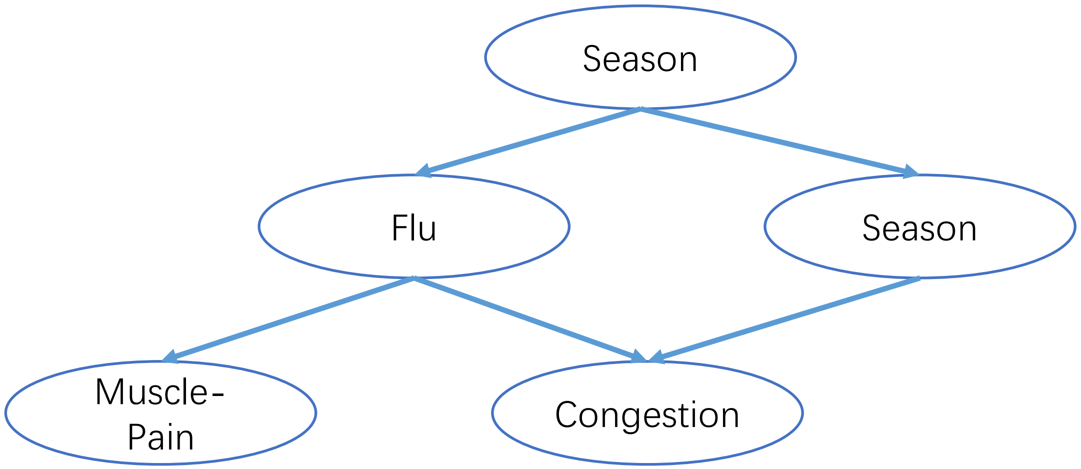
概率图模型

用途：简化联合概率的计算

用自由度
表示需要计算的概率值个数



概率图模型



计算P(SFHMS)需要计算63个概率值
S取值有4种,
F,H,M,C取值各2种,
故自由度=4*2*2*2*2-1=63

由概率图进行因子分解则
P(S,F,H,C,M)
=P(M|SHFC)P(SHFC)
=P(M|F)P(SHFC)
=P(M|F)P(C|SFH)P(SFH)
=P(M|F)P(C|FH)P(SFH)
=P(M|F)P(C|FH)P(F|SH)P(SH)
=P(M|F)P(C|FH)P(F|S)P(H|S)P(S)

P(S)自由度=4,
P(F|S)自由度=3,
P(F|S) P(y|Spring),P(n|Spring) 自由度=1
P(y|Summer),P(n|Summer) 自由度=1
P(y|Autumn),P(n|Autumn) 自由度=1
P(y|Winter),P(n|Winter) 自由度=1
P(M|F) 自由度2
P(C|FH) 自由度4
P(F|S) 自由度4
2+4+4+4+3=17

产生式模型与判别式模型

	产生式分类模型	判别式分类模型
思路	先求P(X,Y)，再求P(Y|X)	直接求P(Y|X)
模型	朴素贝叶斯、HMM	逻辑回归、CRF

朴素贝叶斯原理概述

$$P(Y|X) = \frac{P(XY)}{P(X)}$$

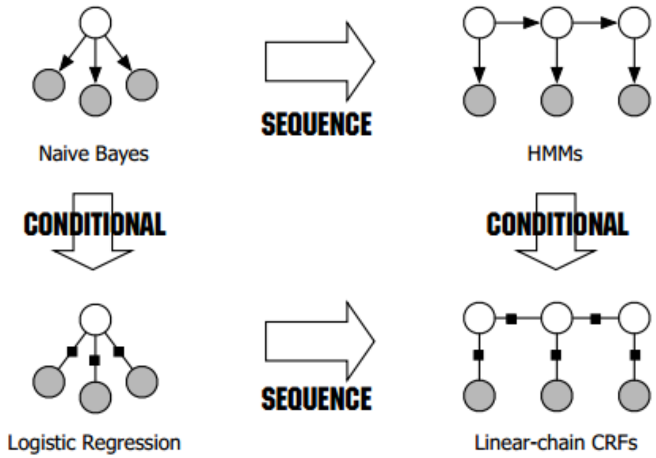
联合概率

$$= \frac{P(X|Y)P(Y)}{P(X)}$$

逻辑回归原理概述

四种模型的概率图模型以及关系

序列化朴素贝叶斯



隐马尔可夫模型的基本概念

- 隐马尔可夫模型是关于时序的概率模型，描述一个隐藏的马尔科夫链随机生成不可观测的状态序列，再由各个状态序列随机生成一个观测而产生观测的序列的过程
- 隐马尔可夫模型的三要素：状态转移概率矩阵A，初始状态分布矩阵π，发射（观测）概率矩阵B
- 隐马尔可夫模型是一个**生成模型**，表示状态序列和观测序列的**联合分布**，但是状态序列是隐藏的，不可观测的
- 隐马尔可夫模型可以用于标注，这时状态对应着标记。标注问题是给定观测序列预测其对应的标记序列
- 马尔科夫模型的两个基本假设
  - 齐次马尔可夫性假设

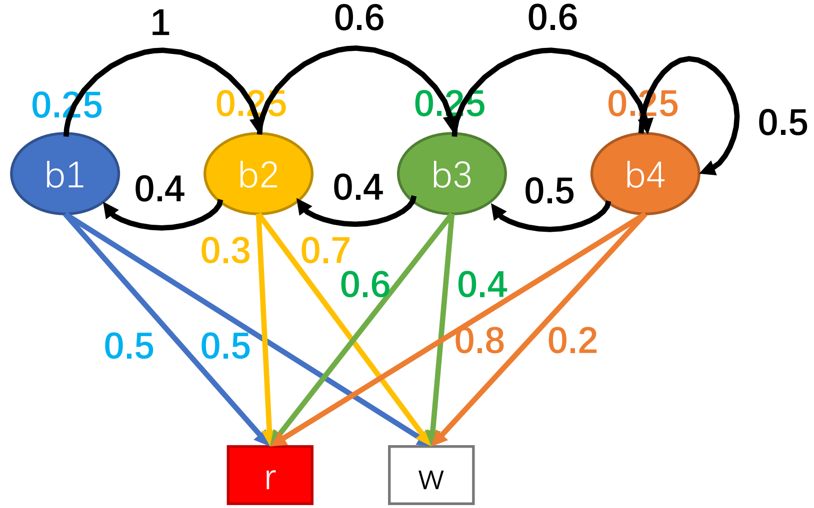
    假设隐藏的马尔可夫链在任意时刻t的状态只依赖于其前一时刻的状态，与其他时刻的状态及观测无关，也与时刻t无关
  - 观测独立性假设

    假设任意时刻的观测只依赖于该时刻的马尔可夫链状态，与其他观测及状态无关

通过一个例子，理解马尔科夫、隐马的相关术语及含义

（1）等概率随机选取1个盒子
（2）从一个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，那么下一个盒子一定是盒子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
（3）每个盒子球的颜色及数量

盒子	1	2	3	4
红球数	5	3	6	8
白球数	5	7	4	2



观测序列 r r w w r

（1）$\pi = [0.25,0.25,0.25,0.25]$ 初始概率分布

（2）$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$ （3）$B = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$

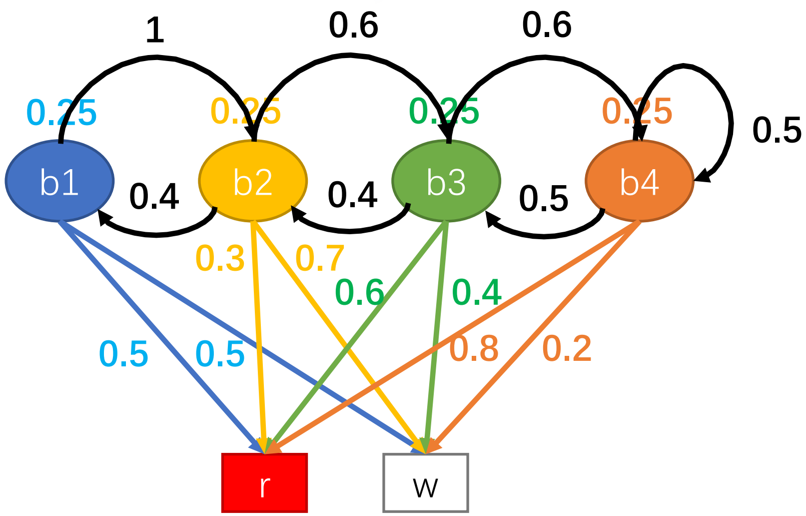
状态转移概率分布 发射矩阵

隐马尔可夫模型的三个基本问题

- 基本问题一
  - 问题描述：已知模型$\mu = (A, B, \pi)$，一个观测序列$O = (O_1, O_2, \dots, O_T)$，求$P(O|\mu)$
  - 计算方法：动态规划——前向、后向算法
- 基本问题二
  - 问题描述：已知模型$\mu = (A, B, \pi)$，一个观测序列$O = (O_1, O_2, \dots, O_T)$，求$P(Q|\mu, O)$
  - 计算方法：动态规划——维特比算法
- 基本问题三
  - 问题描述：已知一个观测序列$O = (O_1, O_2, \dots, O_T)$，求参数$\mu$，使得似然函数$P(O|\mu)$最大，即
    $$\arg\max_{\mu} P(O|\mu)$$
  - 计算方法：（1）极大似然估计（2）EM算法——BW算法

基本问题1——概率计算问题（前、后向算法）

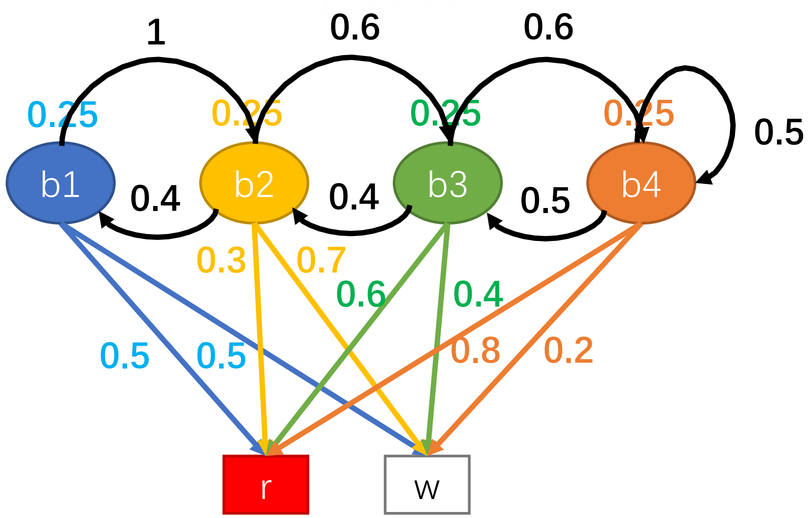
已知　　计算观测到序列概率



观测序列 r r w w r

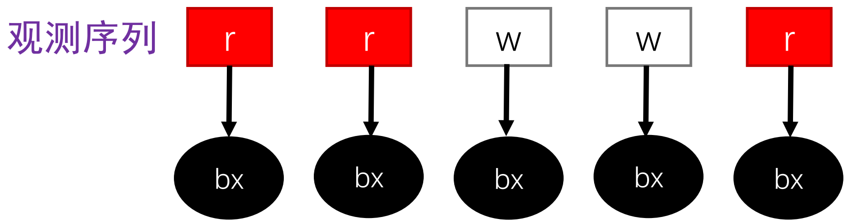
基本问题2——观测/解码问题（Viterbi算法）

已知



观测序列 r r w w r

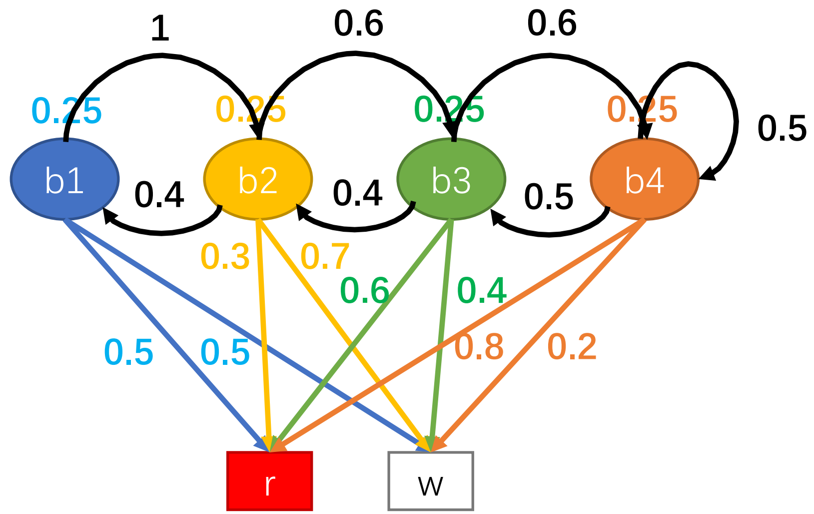
求最佳隐含状态



基本问题3——学习问题（Baum-Welch算法）

已知　　　　估计模型中的数值

观测序列 r r w w r

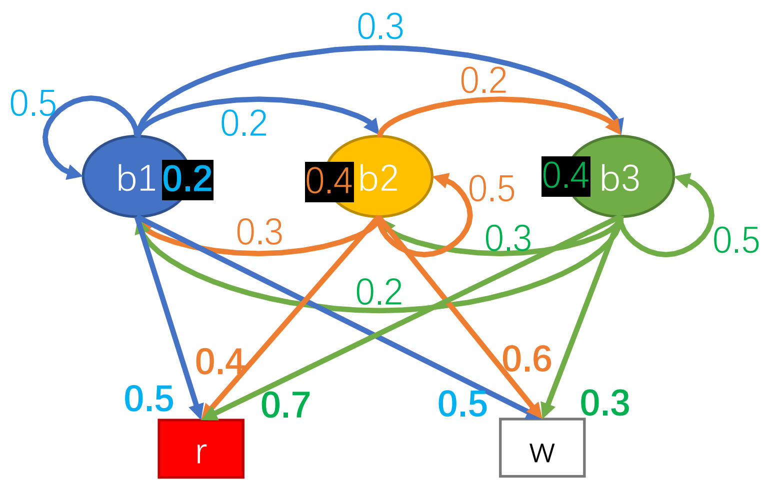


求解基本问题1——前向算法

- 给定盒子和球组成的隐马尔可夫模型$\lambda=(A,B,\pi)$，其中，

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

设T=4，O=(红,白,红,白)，试用前向算法计算$P(O|\lambda)$.



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,⋯, O_t且状态为q_i的概率为前向概率，记作

$$\alpha_{\mathrm{t}}(i) = P(O_1, O_2, \ldots, O_t, i_t = q_i|\lambda)$$

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：观测序列的概率$P(O|\lambda)$.

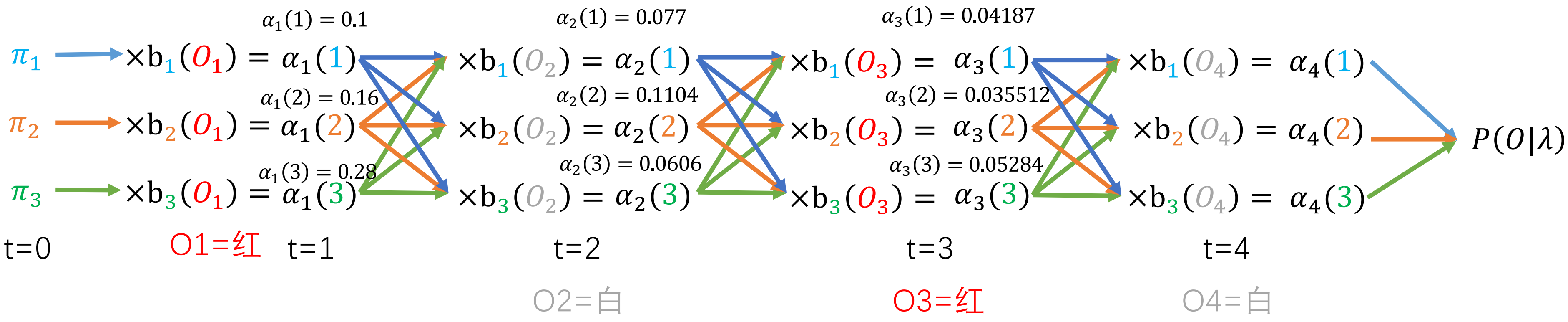
（1）初值 $\alpha_1(i) = \pi_i b_i(O_1)$，i=1,2,⋯,N N表示状态数量

（2）递推 对t=1,2,⋯,T-1,

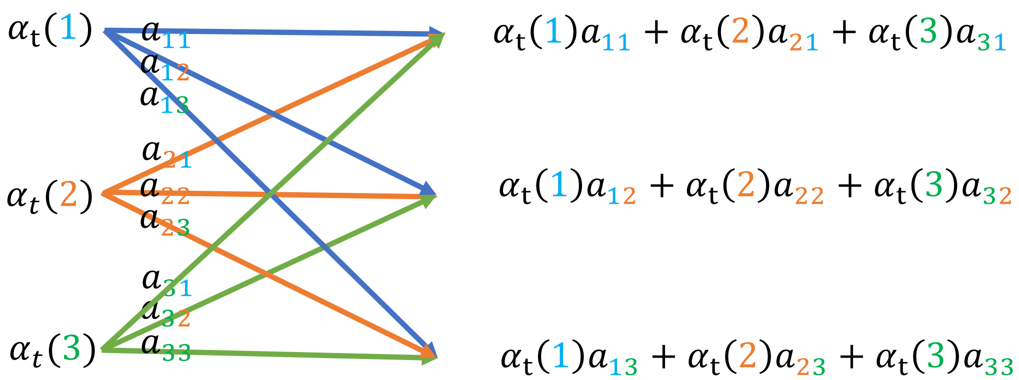
$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i(O_{t+1})$$ ， i=1,2,⋯,N N表示状态数量

（3）终止

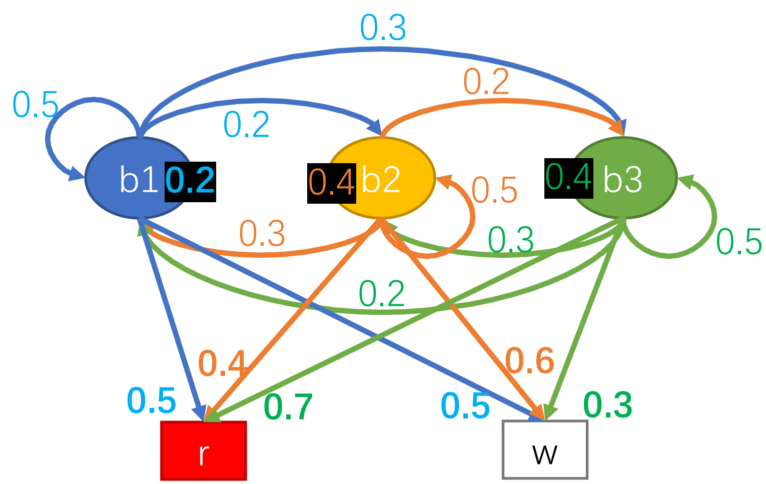
$$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$



α1(1) = 0.1
α1(2) = 0.16
α1(3) = 0.28
α2(1) = 0.077
α2(2) = 0.1104
α2(3) = 0.0606
α3(1) = 0.04187
α3(2) = 0.035512
α3(3) = 0.05284
π1 ×b1(O1) = α1(1) ×b1(O2) = α2(1) ×b1(O3) = α3(1) ×b1(O4) = α4(1)
π2 ×b2(O1) = α1(2) ×b2(O2) = α2(2) ×b2(O3) = α3(2) ×b2(O4) = α4(2) P(O|λ)
π3 ×b3(O1) = α1(3) ×b3(O2) = α2(3) ×b3(O3) = α3(3) ×b3(O4) = α4(3)
t=0
O1=红
t=1
t=2
t=3
t=4
O2=白
O3=红
O4=白



αt(1)
αt(2)
αt(3)
a11
a12
a13
a21
a22
a23
a31
a32
a33
αt(1)a11 + αt(2)a21 + αt(3)a31
αt(1)a12 + αt(2)a22 + αt(3)a32
αt(1)a13 + αt(2)a23 + αt(3)a33



0.3
0.2
0.5
0.2
b1
0.2
0.4
b2
0.4
b3
0.5
0.3
0.3
0.5
0.2
0.4
0.6
0.5
0.7
0.5
0.3
r
w

前向算法

$$\alpha_1(i) = \pi_i b_i(O_1)$$
$$\alpha_1(1) = \pi_1 b_1(O_1) = 0.2\times 0.5=0.1$$
$$\alpha_1(2) = \pi_2 b_2(O_1) = 0.4\times 0.4=0.16$$
$$\alpha_1(3) = \pi_3 b_3(O_1) = 0.4\times 0.7=0.28$$

$$A=\begin{bmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{bmatrix}$$

$$B = \begin{array}{c c} & \begin{array}{cc} \text{v1-红} & \text{v2-白} \end{array} \\ \begin{array}{c} \text{b1} \\ \text{b2} \\ \text{b3} \end{array} & \begin{bmatrix} 0.5 & 0.5 \\ 0.4 & 0.6 \\ 0.7 & 0.3 \end{bmatrix} \end{array}$$

$$\pi = [0.2,0.4,0.4]$$

$$\alpha_2(j) = \left(\sum_{i=1}^{3} \alpha_1(i) a_{ij}\right) b_j(O_2)$$

$$\alpha_2(1) = \left(\sum_{i=1}^{3} \alpha_1(i) a_{i1}\right) b_1(O_2) = (\alpha_1(1)a_{11} + \alpha_1(2)a_{21} + \alpha_1(3)a_{31}) b_1(O_2)$$
$$= (0.1\times 0.5 + 0.16\times 0.3 + 0.28\times 0.2)\times 0.5=0.077$$
$$\alpha_2(2) = \left(\sum_{i=1}^{3} \alpha_1(i) a_{i2}\right) b_2(O_2) = (\alpha_1(1)a_{12} + \alpha_1(2)a_{22} + \alpha_1(3)a_{32}) b_2(O_2)$$
$$= (0.1\times 0.2 + 0.16\times 0.5 + 0.28\times 0.3)\times 0.6=0.1104$$
$$\alpha_2(3) = \left(\sum_{i=1}^{3} \alpha_1(i) a_{i3}\right) b_3(O_2) = (\alpha_1(1)a_{13} + \alpha_1(2)a_{23} + \alpha_1(3)a_{33}) b_3(O_2)$$
$$= (0.1\times 0.3 + 0.16\times 0.2 + 0.28\times 0.5)\times 0.3=0.0606$$

$$\alpha_3(j) = \left(\sum_{i=1}^{3} \alpha_2(i) a_{ij}\right) b_j(O_3)$$

$$\alpha_3(1) = \left(\sum_{i=1}^{3} \alpha_2(i) a_{i1}\right) b_1(O_3) = (\alpha_2(1)a_{11} + \alpha_2(2)a_{21} + \alpha_2(3)a_{31}) b_1(O_3)$$
$$= (0.077\times 0.5 + 0.1104\times 0.3 + 0.0606\times 0.2)\times 0.5 = 0.04187$$
$$\alpha_3(2) = \left(\sum_{i=1}^{3} \alpha_2(i) a_{i2}\right) b_2(O_3) = (\alpha_2(1)a_{12} + \alpha_2(2)a_{22} + \alpha_2(3)a_{32}) b_2(O_3)$$
$$= (0.077\times 0.2 + 0.1104\times 0.5 + 0.0606\times 0.3)\times 0.4 = 0.035512$$
$$\alpha_3(3) = \left(\sum_{i=1}^{3} \alpha_2(i) a_{i3}\right) b_3(O_3) = (\alpha_2(1)a_{13} + \alpha_2(2)a_{23} + \alpha_2(3)a_{33}) b_3(O_3)$$
$$= (0.077\times 0.3 + 0.1104\times 0.2 + 0.0606\times 0.5)\times 0.7 = 0.05284$$

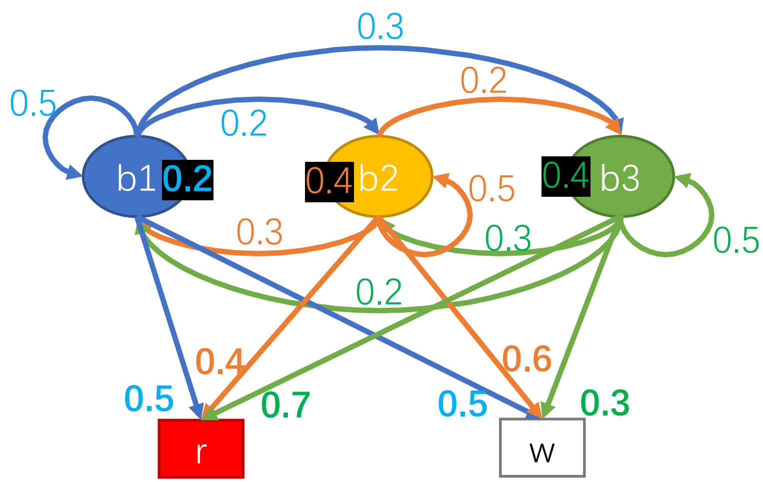
$$P(O|\mu) = \sum_{i=1}^{3} \alpha_3(i) = \alpha_3(1)+\alpha_3(2)+\alpha_3(3)$$
$$= 0.04187+0.035512+0.05284=0.13022$$

求解基本问题1——后向算法

■ 给定盒子和球组成的隐马尔可夫模型$\lambda=(A,B,\pi)$，其中，

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

设T=4，O=(红,白,红,白)，试用后向算法计算$P(O|\lambda)$.



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}，…，o_{T}的概率为后向概率，记作

$$\beta_t(i) = P(\mathrm{o}_{t+1}，\mathrm{o}_{t+2}，…，\mathrm{o}_{\mathrm{T}} | i_t = q_i, \lambda)$$

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：观测序列的概率$P(O|\lambda)$.

（1）初值 $\beta_{\mathrm{T}}(i) = 1$，i=1,2,⋯,N N表示状态数量

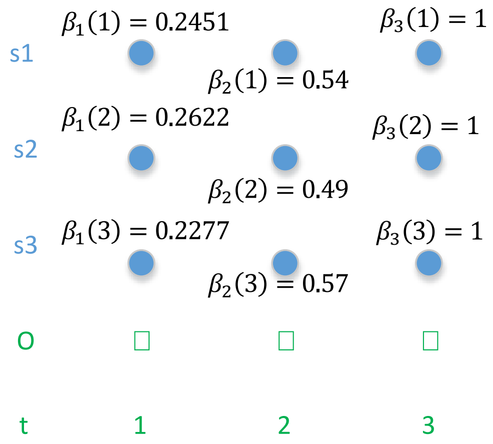
（2）递推 对t=T-1,T-2,⋯,1,

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j)，$$ i=1,2,⋯,N N表示状态数量

（3）终止

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

后向算法



$$\beta_3(1)=\beta_3(2) = \beta_3(3) = 1$$

$$\beta_2(i) = \sum_{j=1}^{3} a_{ij} b_j(O_3)\beta_3(j)$$

$\beta_2(1) = \sum_{j=1}^{3} a_{1j} b_j(O_3)\beta_3(j) = a_{11}b_1(O_3)\beta_3(1)+a_{12}b_2(O_3)\beta_3(2) + a_{13}b_3(O_3)\beta_3(3)$

$=0.5\times0.5\times1 + 0.2\times0.4\times1 + 0.3\times0.7\times1 = 0.54$

$\beta_2(2) = \sum_{j=1}^{3} a_{2j} b_j(O_3)\beta_3(j) = a_{21}b_1(O_3)\beta_3(1)+a_{22}b_2(O_3)\beta_3(2) + a_{23}b_3(O_3)\beta_3(3)$

$=0.3\times0.5\times1 + 0.5\times0.4\times1 + 0.2\times0.7\times1 = 0.49$

$\beta_2(3) = \sum_{j=1}^{3} a_{3j} b_j(O_3)\beta_3(j) = a_{31}b_1(O_3)\beta_3(1)+a_{32}b_2(O_3)\beta_3(2) + a_{33}b_3(O_3)\beta_3(3)$

$=0.2\times0.5\times1 + 0.3\times0.4\times1 + 0.5\times0.7\times1 = 0.57$

$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix}$$

$$\beta_1(i) = \sum_{j=1}^{3} a_{ij} b_j(O_2)\beta_2(j)$$

v1-红 v2-白

$$B = \begin{array}{c}b1\\b2\\b3\end{array}\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix}$$

$\beta_1(1) = \sum_{j=1}^{3} a_{1j} b_j(O_2)\beta_2(j) = a_{11}b_1(O_2)\beta_2(1)+a_{12}b_2(O_2)\beta_2(2) + a_{13}b_3(O_2)\beta_2(3)$

$=0.5\times0.5\times0.54 + 0.2\times0.6\times0.49 + 0.3\times0.3\times0.57 = 0.2451$

$\beta_1(2) = \sum_{j=1}^{3} a_{2j} b_j(O_2)\beta_2(j) = a_{21}b_1(O_2)\beta_2(1)+a_{22}b_2(O_2)\beta_2(2) + a_{23}b_3(O_2)\beta_2(3)$

$=0.3\times0.5\times0.54 + 0.5\times0.6\times0.49 + 0.2\times0.3\times0.57 = 0.2622$

$\beta_1(3) = \sum_{j=1}^{3} a_{3j} b_j(O_2)\beta_2(j) = a_{31}b_1(O_2)\beta_2(1)+a_{32}b_2(O_2)\beta_2(2) +$
$a_{33}b_3(O_2)\beta_2(3)$

$\pi = [0.2,0.4,0.4]$

$=0.2\times0.5\times0.54 + 0.3\times0.6\times0.49 + 0.5\times0.3\times0.57 = 0.2277$

$\mathrm{P}(O|\mu) = \sum_{i=1}^{3} \pi_i\beta_1(i) = \pi_1\beta_1(1)+\pi_2\beta_1(2)+\pi_3\beta_1(3)$

$=0.2\times0.5\times0.2451 + 0.4\times0.4\times0.2622 + 0.4\times0.7\times0.2277 = 0.130218$

求解基本问题2——维特比算法

- 维特比算法是用动态规划解隐马尔可夫模型预测问题，即用动态规划（dynamic programming）求概率最大路径（最优路径）。这时一条路径对应着一个状态序列。
- 根据动态规划原理，最优路径具有这样的特性：如果最优路径在时刻t通过结点i_t^*，那么这一路径从结点i_t^*到终点i_T^*的部分路径，对于从i_t^*到i_T^*的所有可能的部分路径来说，必须是最优的。
- 依据上面的这一原理，我们只需从时刻t=1开始，递推地计算在时刻t状态为i的各条部分路径的最大概率，直至得到时刻t=T状态为i的各条路径的最大概率

求解基本问题2——维特比算法

- 时刻t=T的最大概率即为最优路径的概率P^*，最优路径的终结点i_{T}^*也同时得到
- 之后，为了找出最优路径的各个结点，从终结点i_{T}^*开始，由后向前逐步求得结点$i_{\mathrm{T}-1}^*$，$i_{\mathrm{T}-2}^*$，⋯，i_1^*，得到最优路径$\mathrm{I}^* = (\mathrm{i}_1^*, \mathrm{i}_2^*, \dots, \mathrm{i}_T^*)$

定义 在时刻t状态为i的所有单个路径$(\mathrm{i}_1, \mathrm{i}_2, \ldots, \mathrm{i}_t)$中概率最大值为

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

δ的递推公式：

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

$$= \max_{1 \le j \le N} \left[\delta_t(j) a_{ji}\right] b_i(o_{t+1}), \quad i = 1,2, \ldots, N; \ \mathrm{t}=1,2,\cdots,\mathrm{T}-1$$

定义 在时刻t状态为i的所有单个路径$(\mathrm{i}_1, \mathrm{i}_2, \ldots, \mathrm{i}_{t-1}, i_t)$中概率最大的路径的第$t-1$个结点为

$$\Psi_{\mathrm{t}}(i) = arg \max_{1 \le j \le N} \left[\delta_{t-1}(j) a_{ji}\right], \quad i = 1,2, \ldots, N$$

维特比算法算法

输入：隐马尔可夫模型λ，观测序列O；

输出：最优路径$\mathrm{I}^* = (i_1^*, i_2^*, \dots, i_T^*)$

（1）初始化

$$\delta_1(\mathrm{i}) = \pi_i b_i(o_1),\quad i = 1,2,\dots,N$$

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

（2）递推 对t=2,3,⋯,T

$$\delta_t(\mathrm{i}) = \max_{1\le j\le N}\left[\delta_{t-1}(j)a_{ji}\right]b_i(o_t),\quad i = 1,2,\dots,N$$

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

（3）终止

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

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

（4）最优路径回溯 对t=T-1,T-2,..,1

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

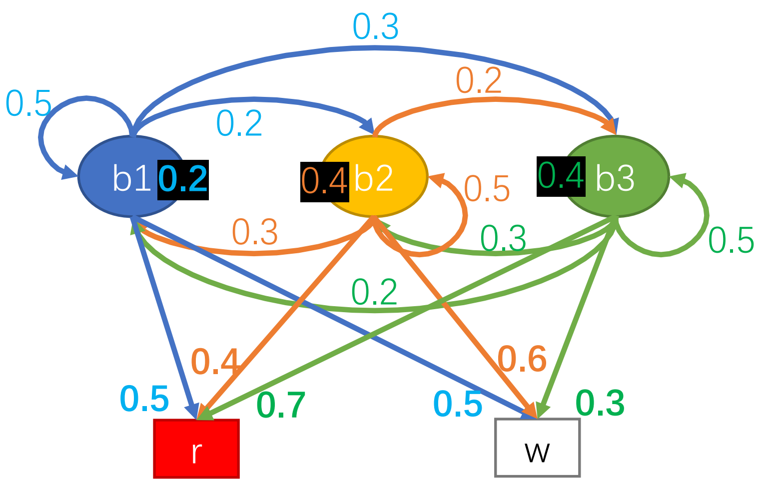
求得最优路径$\mathrm{I}^* = (\mathrm{i}_1^*, \mathrm{i}_2^*, \dots, \mathrm{i}_T^*)$

求解基本问题2——维特比算法

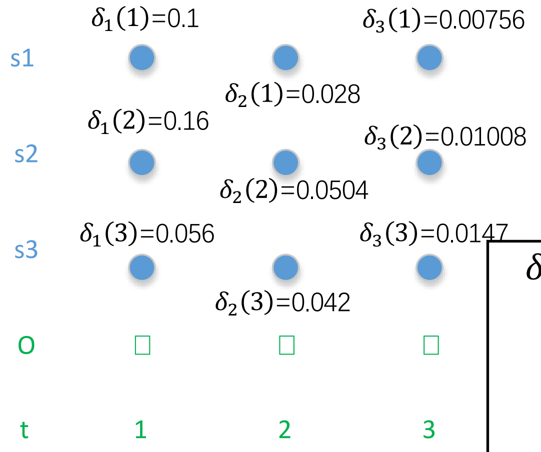
■ 给定盒子和球组成的隐马尔可夫模型$\lambda=(A,B,\pi)$，其中，

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

设T=4，O=(红,白,红)，试用维特比算法求解最优路径$\mathrm{I}^*=(\mathrm{i}_1^*,\mathrm{i}_2^*,\mathrm{i}_3^*)$.



维特比算法



$$A=\begin{bmatrix}0.5 & 0.2 & 0.3\\0.3 & 0.5 & 0.2\\0.2 & 0.3 & 0.5\end{bmatrix}$$

v1-红 v2-白

$$B\ \begin{array}{c}\text{b1}\\\text{b2}\\\text{b3}\end{array}\begin{bmatrix}0.5 & 0.5\\0.4 & 0.6\\0.7 & 0.3\end{bmatrix}$$

π =[0.2,0.4,0.4]

$$\delta_1(\mathrm{i}) = \pi_i b_i(O_1)$$
$$\delta_1(1) = \pi_1 b_1(O_1) = 0.2\times0.5 = 0.1$$
$$\delta_1(2) = \pi_2 b_2(O_1) = 0.4\times0.4 = 0.16$$
$$\delta_1(3) = \pi_3 b_3(O_1) = 0.4\times0.7 = 0.28$$
$$\psi_1(1) = \psi_1(2) = \psi_1(3) = 0$$

$$\delta_2(j) = \max_{1\le i\le 3}\left[\delta_1(i)a_{ij}\right] b_j(O_2)$$

$$\delta_2(1) = \max_{1\le i\le 3}[\delta_1(i)a_{i1}]\, b_1(O_2)$$
$$=0.056\times0.5 = 0.028$$
$$\delta_1(1)a_{11}=0.1\times0.5 = 0.05$$
$$\delta_1(2)a_{21}=0.16\times0.3 = 0.048$$
$$\delta_1(3)a_{31}=0.28\times0.2 = 0.056 \quad \psi_2(1) = 3$$

$$\delta_2(2) = \max_{1\le i\le 3}[\delta_1(i)a_{i2}]\, b_2(O_2)$$
$$=0.084\times0.6 = 0.0504$$
$$\delta_1(1)a_{12} = 0.1\times0.2 = 0.02$$
$$\delta_1(2)a_{22} = 0.16\times0.5 = 0.08$$
$$\delta_1(3)a_{32} = 0.28\times0.3 = 0.084 \quad \psi_2(2) = 3$$

$$\delta_2(3) = \max_{1\le i\le 3}[\delta_1(i)a_{i3}]\, b_3(O_2)$$
$$=0.14\times0.3 = 0.042$$
$$\delta_1(1)a_{13} = 0.1\times0.3 = 0.03$$
$$\delta_1(2)a_{23} = 0.16\times0.2 = 0.032$$
$$\delta_1(3)a_{33} = 0.28\times0.5 = 0.14 \quad \psi_2(3) = 3$$

$$\delta_3(j) = \max_{1\le i\le 3}\left[\delta_2(i)a_{ij}\right] b_j(O_3)$$

$$\delta_3(1) = \max_{1\le i\le 3}[\delta_2(i)a_{i1}]\, b_1(O_3)$$
$$=0.01512\times0.5 = 0.00756$$
$$\delta_2(1)a_{11} = 0.028\times0.5 = 0.014$$
$$\delta_2(2)a_{21} = 0.0504\times0.3 = 0.015 \quad \psi_3(1) = 2$$
$$\delta_2(3)a_{31} = 0.042\times0.2 = 0.0084$$

$$\delta_3(2) = \max_{1\le i\le 3}[\delta_2(i)a_{i2}]\, b_2(O_3)$$
$$=0.0252\times0.4 = 0.01008$$
$$\delta_2(1)a_{12} = 0.028\times0.2 = 0.0056$$
$$\delta_2(2)a_{22} = 0.0504\times0.5 = 0.0252 \quad \psi_3(2) = 2$$
$$\delta_2(3)a_{32} = 0.042\times0.3 = 0.0126$$

$$\delta_3(3) = \max_{1\le i\le 3}[\delta_2(i)a_{i3}]\, b_3(O_3)$$
$$=0.021\times0.7 = 0.0147$$
$$\delta_2(1)a_{13} = 0.028\times0.3 = 0.0084$$
$$\delta_2(2)a_{23} = 0.0504\times0.2 = 0.01008$$
$$\delta_2(3)a_{33} = 0.042\times0.5 = 0.021 \quad \psi_3(3) = 3$$

$$\widehat{q_3} = \arg\max_{0\le i\le 3}[\delta_3(i)] = 3$$
$$\widehat{q_t} = \psi_{t+1}(\widehat{q_{t+1}})$$
$$\widehat{q_2} = \psi_3(\widehat{q_3}) = \psi_3(3) = 3$$
$$\widehat{q_1} = \psi_2(\widehat{q_2}) = \psi_2(3) = 3$$

求解基本问题3——极大似然估计（统计次数方法）

- 在训练语料给出了观测序列$O=(o_1,o_2,\dots,o_T)$以及对应的隐藏状态序列$\mathrm{Q}=(q_1,q_2,\dots,q_T)$的情况下，可以采用最大似然估计法估计模型中的参数$\mu=(A,B,\pi)$

$$\bar{\pi}_i=\delta(q_1,s_i)$$

$$\overline{a_{ij}}=\frac{Q\text{中从状态}q_i\text{转移到}\mathrm{q}_j\text{的次数}}{\mathrm{Q}\text{中从状态}q_i\text{转移到另一状态（包括}q_i\text{自身）的次数}}$$

$$\overline{b_j(k)}=\frac{Q\text{中从状态}\mathrm{q}_j\text{输出符号}v_k\text{的次数}}{Q\text{到达}\mathrm{q}_j\text{的次数}}$$

求解基本问题3——EM算法（Baum-Welch算法）

- 步骤1 初始化

  随机地给参数π_i，a_{ij}，$b_j(k)$赋值，使其满足如下约束

$$\sum_{i=1}^{N}\pi_i = 1 \qquad \sum_{j=1}^{N}a_{ij} = 1, 1 \le i \le N \qquad \sum_{k=1}^{M}b_j(k) = 1, 1 \le j \le N$$

- 步骤2 EM计算
  - E步骤 由模型参数计算期望值

  （1）给定HMM参数μ和观察序列$O=(O_1,O_2,\dots,O_T)$，在时间t位于状态s_i，时间$t+1$位于状态s_j概率

$$\xi_t(i,j)=\frac{\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)}$$

  （2）给定HMM参数μ和观察序列$O=(O_1,O_2,\dots,O_T)$，在时间t位于状态s_i的概率

$$\gamma_t(i)=\sum_{j=1}^{N}\xi_t(i,j)$$

  - M步骤 由期望值重新估计模型参数

$$\overline{\pi_i} = \gamma_t(i) \qquad \overline{a_{ij}} = \frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \qquad \overline{b_j(k)} = \frac{\sum_{t=1}^{T-1}\gamma_t(j)\times\delta(O_t,v_k)}{\sum_{t=1}^{T-1}\gamma_t(j)}$$

求解基本问题3——EM算法（Baum-Welch算法）

- 步骤3 循环计算

  令$i=i+1$，重复执行EM计算，直到π_i，$\mathrm{a_{ij}}$，$\mathrm{b}_j(k)$收敛

求解基本问题3——EM算法（Baum-Welch算法）

- 实际应用中的问题

1. 系统不知道各隐含状态的实际意义
   用训练出的模型实际测试任务，根据状态序列人工判别各隐含状态标签的实际意义
2. 系统不知道模型的具体任务
   根据任务的不同，设置针对特定任务的初始模型参数值

隐马应用——分词

- 观测序列

  小明硕士毕业于中国科学院计算所

- 状态序列
  - 状态说明

    B代表该字是词语中的起始字

    M代表是词语中的中间字

    E代表是词语中的结束字

    S则代表是单字成词
  - 状态序列

    BEBEBMEBEBMEBES

- 分词结果

  小明/硕士/毕业于/中国/科学院/计算/所

隐马应用——分词

- 模型训练

  采用极大似然估计法，估计模型参数$\mu = (A, B, \pi)$
- 测试分词

  采用viterbi算法，计算最优状态序列

参考文献

- 《统计自然语言处理》宗成庆 第6章 隐马尔可夫模型
- 《统计学习方法》 李航 第10章 隐马尔可夫模型
- 分词训练语料 https://github.com/ldanduo/HMM