隐马尔可夫模型(HMM)

洪青阳 副教授

厦门大学信息科学与技术学院
qyhong@xmu.edu.cn

主要内容

1. 基于统计模型框架的识别法
2. 隐马尔可夫模型(HMM)的概念
   (HMM: Hidden Markov Models)
3. HMM的三个基本问题
4. GMM-HMM

1 基于统计模型框架的识别法

1.1 预备知识

(1)条件概率 $P(A|B)$

$$P(A|B) = P(A,B)/P(B)$$

$P(A,B)$：表示A与B的联合概率。

(2)Bayes定理

$$P(A|B) = P(B|A)P(A)/P(B)$$

(3)事件的独立性

$$P(A,B) = P(A)P(B)$$

1.2 基于统计模型框架的识别法

基于统计模型框架的识别法(1)

- 语音识别问题的形式化描述

-设：(1)待识别语音的特征模式：$\boldsymbol{O} = o_1, o_2, \ldots, o_T$

(2)词汇表中第 **n** 个单词：$W(n)$， $1 \leqslant n \leqslant N$

(3)当$\boldsymbol{O}$ 被观察到后，与$\boldsymbol{O}$ 对应的发声内容是单词 $W(n)$的概率：$P(W(n) \mid \boldsymbol{O})$

-语音识别问题的形式化描述：

$$k = \underset{n}{\operatorname{argmax}} \{P(W(n) \mid \boldsymbol{O})\}$$

基于统计模型框架的识别法(2)

- 声学模型与语言模型

$$P(W(n)\,|\,O) = P(O\,|\,W(n))P(W(n))/P(O)$$

$$k = \underset{n}{\operatorname{argmax}}\{P(O\,|\,W(n))\cdot P(W(n))\}$$

声学模型

语言模型

- 模式匹配与统计模型(O :待识语音)

模式匹配		统计模型	
词 汇 表	$W(k),\ 1\leqslant k\leqslant N$	词 汇 表	$W(k),\ 1\leqslant k\leqslant N$
参考模式	$R(k),\ 1\leqslant k\leqslant N$	参考模型	$M(k),\ 1\leqslant k\leqslant N$
失真测度	$D_k = D(O, R(k))$ $-D_k$:DTW距离	概率测度	$P(O\|M(k))$ $-P$: 由$M(k)$生成O的概率
判 别	$n = \underset{1\leqslant k\leqslant N}{\operatorname{argmin}}\{D_k\}$	判 别	$n=\underset{1\leqslant k\leqslant N}{\operatorname{argmax}}\{P(O\|M(k))\}$
识别结果	$W(n)$	识别结果	$W(n)$

2 隐马尔可夫模型(HMM)的概念

2.1 马尔可夫过程

2.2 隐马尔可夫模型的概念

2.3 HMM的要素及其模型描述

2.4 基于HMM的观察符号序列的生成方式

马尔可夫过程

- 语言的马尔可夫模型

  $P(C_i, C_j) = P(C_i)P(C_j|C_i)$

  $P(C_i, C_j, C_k)$
  $= P(C_i)P(C_j|C_i)P(C_k|C_j)$

- 天气的马尔可夫模型

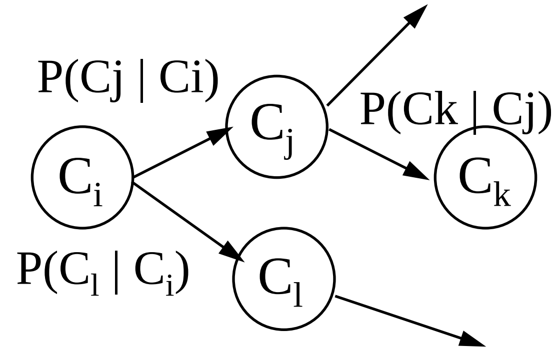
  观察日期：　1 2 3 4 5 6 7 　8

  观察序列(O)：　晴晴晴雨雨晴多云晴

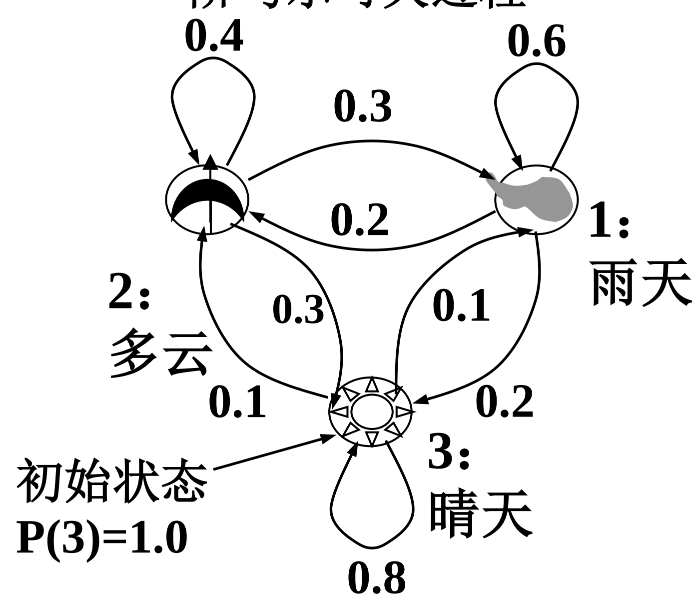
  状态转移序列：3 3 3 1 1 3 2 　3

状态输出概率$P(O|\lambda)$：

$$P(O|\lambda) = P(3,3,3,1,1,3,2,3\ |\lambda)$$
$$= P(3)P(3|3)P(3|3)P(1|3)$$
$$P(1|1)P(3|1)P(2|3)P(3|2)$$



一阶马尔可夫过程



天气的马尔可夫模型

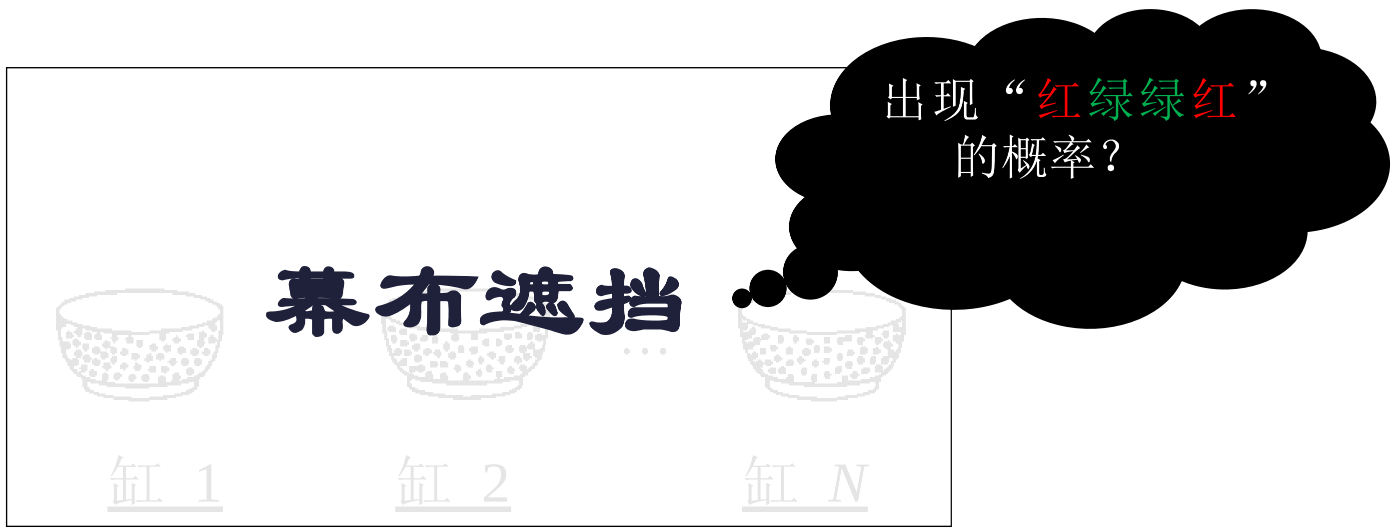
隐马尔可夫模型的概念

隐马尔可夫模型(HMM, Hidden Markov Model)源于Markov链，所描述的问题比Markov模型更复杂。HMM中存在一个双重随机过程：基本的随机过程是状态随机转移；另一个随机过程是观察事件与状态并非一一对应，而是通过一组概率分布相联系，即只能看到观察值，但不能确定观察值对应的状态。

- 双重随机过程
  - 状态转移的随机性
  - 依存于状态的观察事件的随机性

双重随机过程：颜色球例子

假设有N个缸，每个缸中按不同比例存放红绿两种颜色的球。根据初始概率分布，随机地选择其中的一个缸，并随机取出一球，记为O_1。把球放回原缸，根据缸之间的转移概率，随机选择下一个缸，再随机取出一球，记为O_2。如此反复，可以得到一个描述球的颜色的序列O_1，O_2 …，可称之为观察值序列。在此过程中，缸之间的转移不确定，一种球的颜色出自哪个缸也不确定。



$P(\text{红})=b_1(1)$ $P(\text{红})=b_2(1)$ $P(\text{红})=b_N(1)$

$P(\text{绿})=b_1(2)$ $P(\text{绿})=b_2(2)$ $P(\text{绿})=b_N(2)$

HMM模型的要素及其模型描述

- 模型要素：

(1)N：模型中的状态数目

(2)M：每个状态可能输出的观察符号的数目

(3)$A = \{a_{ij}\}$：状态转移概率分布

(4)$B = \{b_j(k)\}$：观察符号的概率分布

(5) $\pi = \{\pi_i\}$：初始状态概率分布

- 模型描述：

$$\lambda = (A, \quad B, \quad \pi)$$

基于HMM的观察符号序列的生成方式

- 当给定模型 $\lambda(A, B, \pi)$ 后，就可将该模型看成一个符号生成器(或称信号源)，由它生成观察序列 $O= o_1o_2 \ldots o_T$。其生成过程(也称HMM过程)是:

(1) 初始状态概率分布 π，随机选择一个初始状态 $q_1 = S_i$;

(2) 置 $t = 1$;

(3) 按状态 S_i 的符号概率分布 $b_i(k)$，随机产生一个输出符号 $o_t = V_k$;

(4) 按状态 S_i 的状态转移概率分布 a_{ij}，随机转移至一个新的状态 $q_{t+1} = S_j$

(5) 令 $t = t + 1$，若 $t \leqslant T$，则返回步骤（3），否则结束过程。

3 HMM的三个基本问题及其解法

3.1 HMM三个基本问题

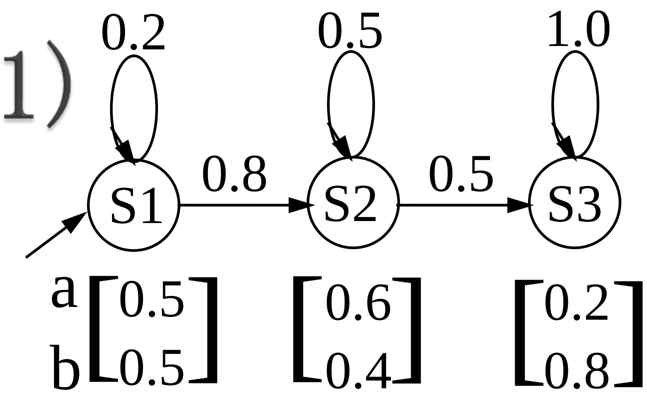
- 模型评估问题(如何求：$P(O|\lambda)$)
- 最佳路径问题(如何求：$Q=q_1q_2\ldots q_T$)
- 模型训练问题(如何求：A、B、π)

3.2 模型评估问题的解法

3.3 最佳路径问题的解法

3.4 模型训练问题的解法

模型评估问题的解法(1)



HMM 模型的例子

观察符号序列：abba

- 当给定模型 $\lambda(A, B, \pi)$ 以及观察序列 $O=o_1o_2\ldots o_T$ 时，计算模型 λ 对观察序列 O 的 $P(O|\lambda)$ 概率的思路是(穷举法)：

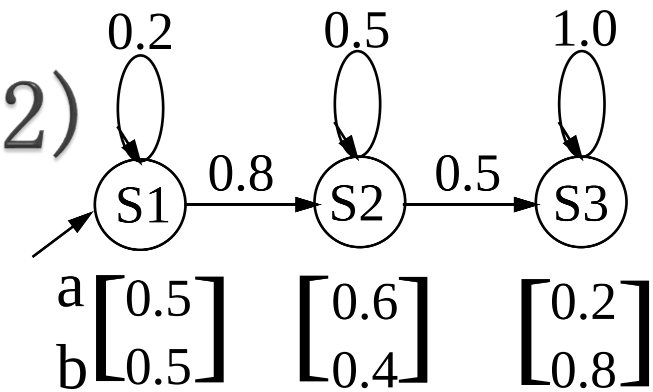
(1)对长度为 T 的观察序列 O，找出所有可能产生该观察序列 O 的状态转移序列 $Q^j=q^j_1\, q^j_2\, q^j_3 \ldots q^j_T\,(j=1, 2, \ldots, J)$；

(2)分别计算 Q^j 与观察序列 O 的联合概率 $P(O, Q^j|\lambda)$；

(2)取各联合概率 $P(O, Q^j|\lambda)$ 的和，即：

$$P(O|\lambda)=\sum_{j=1}^{J} P(O, Q^j|\lambda)$$

模型评估问题的解法(2)



HMM 模型的例子

- $P(O|\lambda)$的一般解法:

$\because\ P(O,Q^j|\lambda) = P(Q^j|\lambda)P(O|Q^j,\lambda)$

$P(Q^j|\lambda) = P(q^j_1)P(q^j_2|q^j_1)P(q^j_3|q^j_2)\ \dots\ P(q^j_T|q^j_{T-1})$

$= a^j_{0,1}\ a^j_{1,2}\ a^j_{2,3}\ \dots a^j_{T-1,T}$

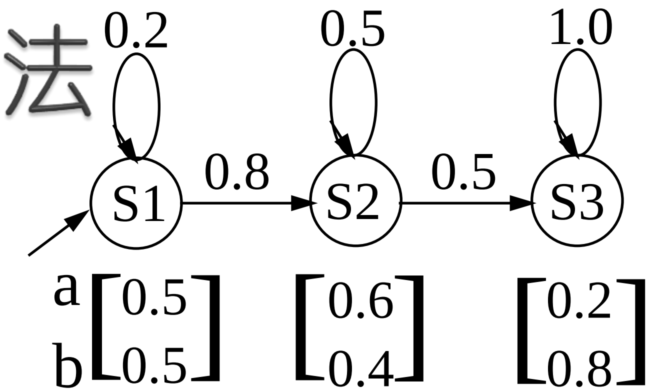
$P(O|Q^j,\lambda) = P(o_1|q^j_1)P(o_2|q^j_2)\ \dots\ P(o_T|q^j_T)$

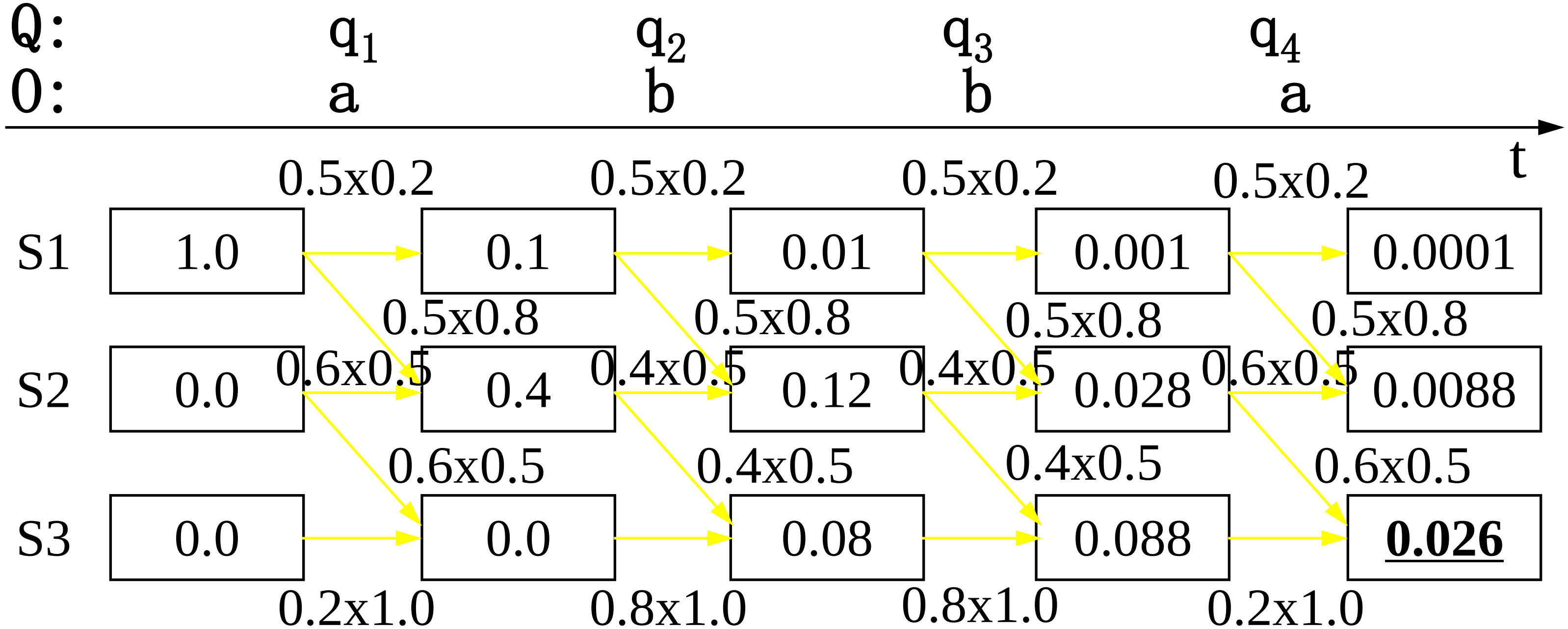
$= b_1{}^j(o_1)\ b_2{}^j(o_2)\ b_3{}^j(o_3)\ \dots\ b_T{}^j(o_T)$

$\therefore\ P(O,Q^j|\lambda) = a^j_{0,1}b_1{}^j(o_1)\ \ a^j_{1,2}\ b_2{}^j(o_2)\ \dots\ a^j_{T-1,T}\ b_T{}^j(o_T)$

$$P(O|\lambda)=\sum_{j=1}^{J}P(O,Q^j|\lambda)=\sum_{j=1}^{J}\left\{\prod_{t=1}^{T}a^j_{t-1,t}b_t{}^j(o_t)\right\}$$

模型评估问题的前向算法





采用前向算法求解P(abba | λ)概率的格型图

前向-后向算法

前向-后向(Forward-Backward)算法用来解决高效计算$P(O|\lambda_i)$ 的问题。该算法分为两部分：

◆ 前向算法

前向算法按输出观察值序列的时间，从前向后地推计算输出概率。此算法用$\alpha_t(j)$表示已经输出部分观察值 $o_1, o_2, ..., o_t$ ，并且到达状态为 θ_j 的概率： $\alpha_t(j) = P(o_1, o_2, ..., o_t, S_t = \theta_j \mid \lambda)$

1. 初始化

$$\alpha_1(i) = \pi_i b_i(\boldsymbol{o}_1), \quad 1 \leq i \leq N$$

2. 迭代计算

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right] b_j(\boldsymbol{o}_{t+1}), \quad \begin{matrix} 1 \leq t \leq T-1 \\ 1 \leq j \leq N \end{matrix}$$

3. 终止计算

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

前向-后向算法

$$o_1, o_2, \dots, o_t o_{t+1}, o_{t+2}, \dots, o_T$$

◆ 后向算法

向后算法由后向前推算输出概率。用$\beta_t(i)$表示时刻t时的状态为θ_i，输出结束时的状态为θ_N，且输出观察值序列为$o_t o_{t+1}, \dots, o_T$的概率：

$$\beta_t(i) = P(o_t, o_{t+1}, \dots, o_T, S_t = \theta_i, S_T = \theta_N \mid \lambda)$$

1. 初始化

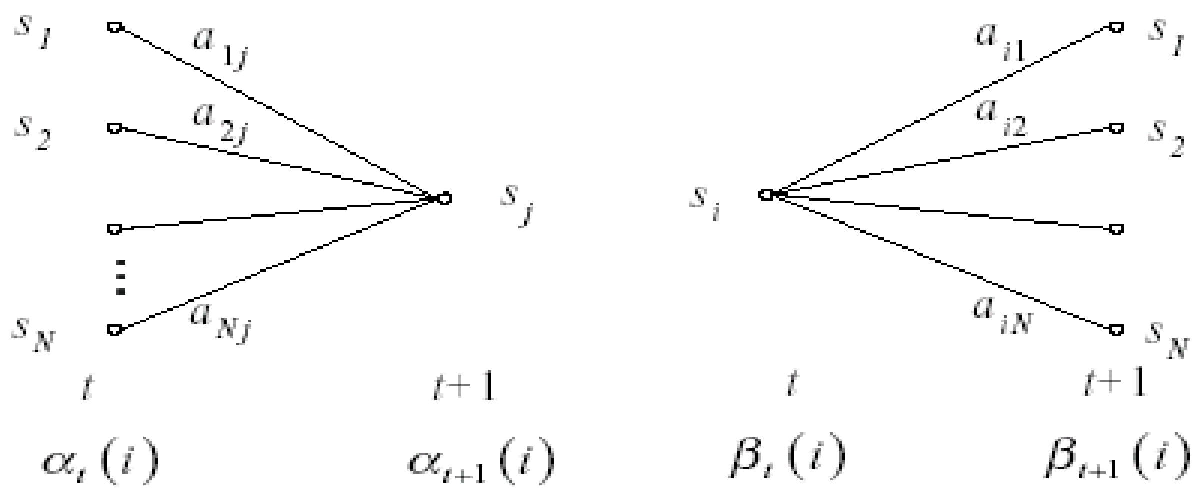
$$\beta_T(i) = 1, \quad 1 \le i \le N$$

2. 迭代计算

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(\boldsymbol{o}_{t+1}) \beta_{t+1}(j), \quad \begin{array}{c} t = T-1, T-2, \cdots, 1 \\ 1 \le j \le N \end{array}$$

前向-后向算法

- 前向和后向递推的示意图



(a) 前向概率的递推　　(b) 后向概率的递推

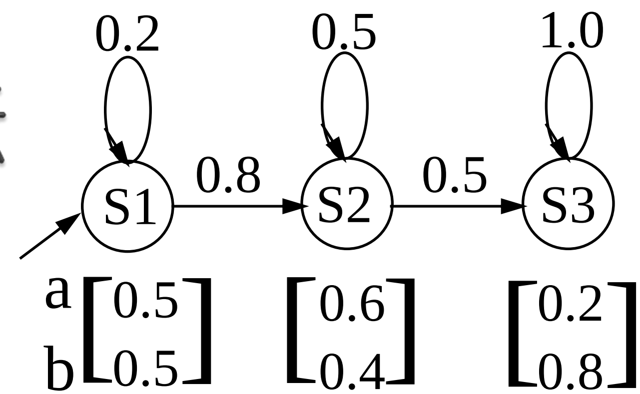
前向-后向算法

- HMM的前向、后向概率估计

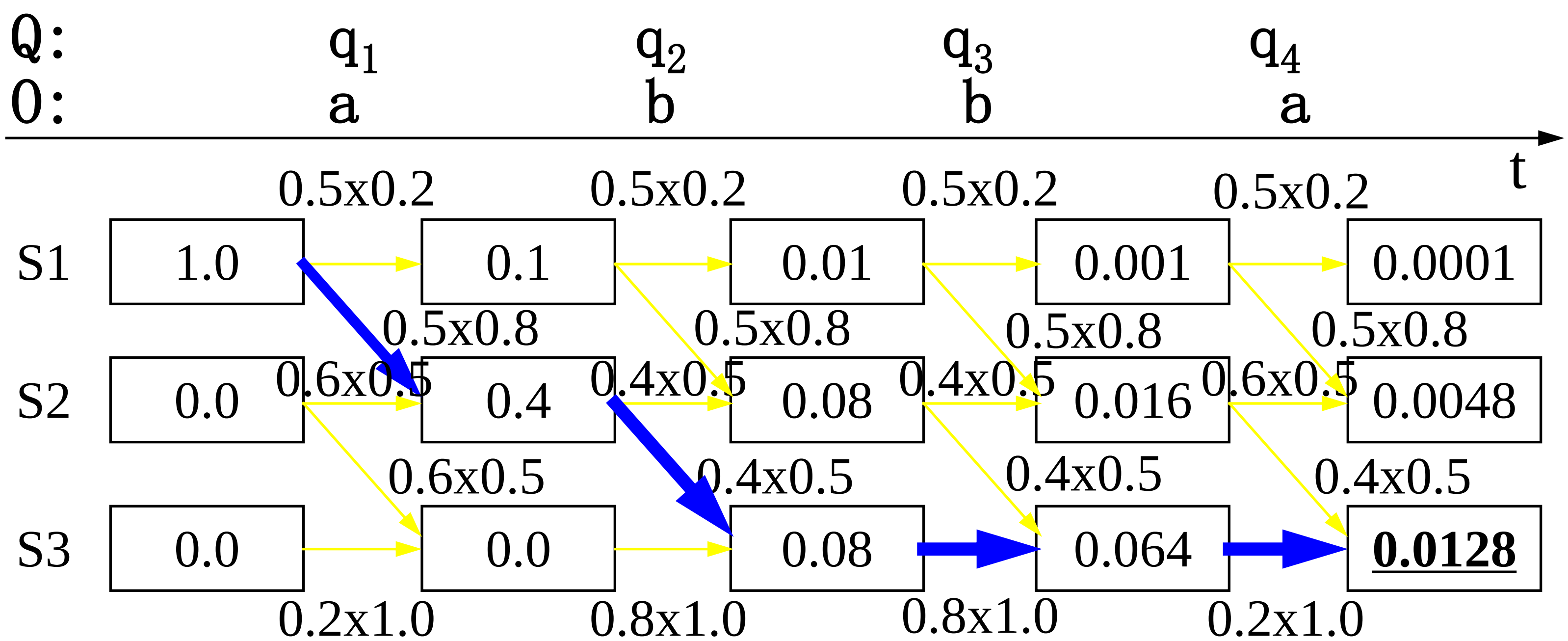
$$P(\boldsymbol{O}|\lambda) = \sum_{i=1}^{N} \alpha_t(i)\beta_t(i) = \sum_{i=1}^{N} \alpha_T(i), \quad 1 \leq t \leq T-1$$

$$P(\boldsymbol{O}|\lambda) = \sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i)a_{ij}b_j(\boldsymbol{o}_{t+1})\beta_{t+1}(j), \quad 1 \leq t \leq T-1$$

最佳路径问题的解法



最佳路径：S1-S2-S3-S3-S3



采用Viterbi算法求解产生观察序列abba最佳路径的格型图

最佳路径问题的解法：Viterbi算法

Viterbi算法用于解决如何寻找与给定观察值序列对应的最佳状态序列的问题。

定义最佳状态序列 $S^* = s_1^*, s_2^*, \ldots, s_T^*$ ，$\varphi_t(j)$记录局部最佳状态序列。

定义$\delta_t(i)$为截止到时刻t，依照状态转移序列$s_1, s_2, \ldots, s_t$，产生出观察值$o_1, o_2, \ldots, o_t$ 的最大概率，且最终状态为θ_i。

$$\delta_t(i) = \max_{s_1, s_2, \ldots s_{t-1}} P(s_1, s_2, \ldots, s_t, S_t = \theta_i, | \lambda)$$

初始化：$\delta_0(1) = 1,\ \delta_0(j) = 0 \quad (j \neq 1)$

$$\varphi_1(j) = S_0$$

递推：$\delta_t(j) = b_{ij}(o_t) \max_{1 \le i \le N} \delta_{t-1}(i) a_{ij} \quad (1 \le t \le T; 1 \le i \le N)$

$$\varphi_t(j) = \arg\max_{1 \le i \le N} \delta_{t-1}(i) a_{ij}$$

最佳路径问题的解法：Viterbi算法

终止：$P_{\max}(S,\mathrm{O}|\lambda)=\max\limits_{1\le i\le N}\delta_T(i)$

$$\varphi_T(N)=\arg\max_{1\le i\le N}\delta_{T-1}(i)a_{ij}$$

算法终止时$\delta_t()$记录的数据便是 $S*$

S 的不同取值使概率值$P(S,\mathrm{O}|\lambda)$ 差别很大，而$P(S^*,\mathrm{O}|\lambda)$是$\sum\limits_s P_{\max}(S,\mathrm{O}|\lambda)$的各个分量中唯一的一个举足轻重的。因此，常等价地用$P(S^*,\mathrm{O}|\lambda)$ 近似 $\sum\limits_s P_{\max}(S,\mathrm{O}|\lambda)$，那么实际上Viterbi算法也就能用来计算 $P(\mathrm{O}|\lambda)$

Viterbi算法的对数形式

0. 预处理

$$\tilde{\pi}_i = \log(\pi_i)$$

$$\tilde{b}_i(o_t) = \log[b_i(o_t)]$$

$$\tilde{a}_{ij} = \log(a_{ij})$$

1. 初始化

$$\tilde{\delta}_1(i) = \log(\delta_1(i)) = \tilde{\pi}_i + \tilde{b}_i(o_t)$$

$$\psi_1(i) = 0 \quad 1 \le i \le N$$

2 递归

$$\tilde{\delta}_t(j) = \log(\delta_t(j)) = \max_{1\le i\le N}\left[\tilde{\delta}_{t-1}(i) + \tilde{a}_{ij}\right] + \tilde{b}_j(o_t)$$

$$\psi_t(j) = \arg\max_{1\le i\le N}\left[\tilde{\delta}_{t-1}(i) + \tilde{a}_{ij}\right] \quad 2 \le t \le T, 1 \le j \le N$$

3. 终止

$$\tilde{P}^* = \max_{1\le i\le N}\left[\tilde{\delta}_T(i)\right]$$

$$q_T^{\ *} = \arg\max_{1\le i\le N}\left[\tilde{\delta}_T(i)\right]$$

4. 回溯

$$q_t^* = \psi_{t+1}(q_{t+1}^*) \quad t = T-1, T-2, \Lambda\ 1$$

HMM参数重估

给定一个观察值序列 $O = o_1, o_2, \Lambda\ o_T$ 确定一个 $\lambda = (\pi, A, B)$，使 $P(O \mid \lambda)$ 最大。

实际上，不存在一种方法直接估计最佳的 λ。

替代的方法：

根据观察值序列选取初始模型 $\lambda = (\pi, A, B)$，然后依据某种方法求得一组新参数 $\bar{\lambda} = (\bar{\pi}, \bar{A}, \bar{B})$，保证有 $P(O \mid \bar{\lambda}) > P(O \mid \lambda)$。重复这个过程，逐步改进模型参数，直到 $P(O \mid \bar{\lambda})$收敛。

HMM参数重估

- 这一方法，未必能求得全局最大值，而有可能得到一局部极值点。
- 经典的方法：Baum-Wel
法）。
- Baum-Welch算法的理论

- EM(Expectation Maximiz
算法

- EM算法是一种从“不完全数据”中求解模型分布参数的最大似然(Maximum Likelihood)估计方法。

其二是对参数的似然函数直接优化十分困难，而引入额外的参数（隐含的）后就比较容易优化。

HMM参数重估

- HMM模型中，数据是由两部分组成，一是可以观测到的，如观测特征序列，称为可观测数据；一是无法观测到，如状态序列，称为隐含变量(Latent Variable)。

- 它们可以共同构成一个完全数据集 (O, Q)。

- EM算法的目的是通过迭代地将完全数据集的对数似然度期望最大化，来实现对可观测数据的对数似然度的最大化。

- 包含两个主要方面：一是求期望(expectation)，用E来表示，一是最大化(maximization)，用M来表示。

HMM参数重估

根据Bayes公式，完全数据集的似然度和可观测数据集的似然度之间存在以下关系：

$$P(O,Q\mid\lambda)=P(Q\mid O,\lambda)P(O\mid\lambda)$$

观测数据的对数似然度可以表示为

$$\log P(O\mid\lambda)=\log P(O,Q\mid\lambda)-\log P(Q\mid O,\lambda)$$

对于两个参数集 λ 和 $\bar{\lambda}$，在已知 O 和 λ 的情况下，在完全数据集上求期望，则 ：

$$E[\log P(O\mid\bar{\lambda})\mid O,\lambda]=E[\log P(O,Q\mid\bar{\lambda})\mid O,\lambda]-E[\log P(Q\mid O,\bar{\lambda})\mid O,\lambda]$$

HMM参数重估

根据随机变量数学期望的定义：

有
$$E[\log P(O|\bar{\lambda})|O,\lambda] = \int \log P(O|\bar{\lambda}) P(Q|O,\lambda) dQ$$
$$= \log P(O|\bar{\lambda})$$
$$= L(O,\bar{\lambda})$$

令

$$Q(\lambda,\bar{\lambda}) = E[\log P(O,Q|\bar{\lambda})|O,\lambda] = \int \log P(O,Q|\bar{\lambda}) P(Q|O,\lambda) dQ$$

$$H(\lambda,\bar{\lambda}) = E[\log P(Q|O,\bar{\lambda})|O,\lambda] = \int \log P(Q|O,\bar{\lambda}) P(Q|O,\lambda) dQ$$

则

$$L(O,\bar{\lambda}) = Q(\lambda,\bar{\lambda}) - H(\lambda,\bar{\lambda})$$

HMM参数重估

$$H(\lambda,\lambda)=\int logP(Q|O,\lambda)P(Q|O,\lambda)dQ$$

$$H(\lambda,\bar{\lambda})=\int logP\left(Q\middle|O,\bar{\lambda}\right)P(Q|O,\lambda)dQ$$

Jensen不等式：

$$H(\lambda,\lambda)-H(\lambda,\bar{\lambda})=\int\left(-log\frac{P(Q|O,\bar{\lambda})}{P(Q|O,\lambda)}\right)P(Q|O,\lambda)dQ$$

$$\geq -log\int\frac{P\left(Q\middle|O,\bar{\lambda}\right)}{P(Q|O,\lambda)}P(Q|O,\lambda)dQ$$

$$= -log\int P\left(Q\middle|O,\bar{\lambda}\right)dQ = 0$$

HMM参数重估

由Jensen不等式可以知道 $H(\lambda,\bar{\lambda}) \le H(\lambda,\lambda)$，因此如果可以保证 $Q(\lambda,\bar{\lambda}) \ge Q(\lambda,\lambda)$，就可以保证不等式 $L(O,\bar{\lambda}) \ge L(O,\lambda)$ 成立。

在语音识别中，由于Q是离散的，因此Q函数可表示为

$$Q(\lambda,\bar{\lambda}) = \sum_{\text{所有}Q} P(O,Q \mid \lambda) \log P(O,Q \mid \bar{\lambda})$$

计算 $\frac{\partial Q(\lambda,\bar{\lambda})}{\partial \bar{\lambda}} = 0$，得到一组求取 $\bar{\lambda}$ 的公式，这一组公式就称为重估（Re-Estimation）公式，它们是Baum-Welch算法的核心内容。

HMM参数重估

$$P(O,Q|\lambda) = \pi_{q_0} \prod_{t=1}^{T} a_{q_{t-1}q_t} b_{q_t}(o_t)$$

$$logP(O,Q|\lambda) = log\pi_{q_0} + \sum_{t=1}^{T} loga_{q_{t-1}q_t} + \sum_{t=1}^{T} logb_{q_t}(o_t)$$

辅助函数 $$Q(\lambda,\bar{\lambda}) = \sum_{i=1}^{N} P\left(O,q_0 = i|\bar{\lambda}\right) log\pi_i + \sum_{j=1}^{N}\sum_{t=1}^{T} P\left(O,q_{t-1} = i,q_t = j|\bar{\lambda}\right) loga_{ij} + \sum_{t=1}^{T} P\left(O,q_t = i|\bar{\lambda}\right) logb_i(o_t)$$

模型训练问题的解法：Baum-Welch算法

Baum-Welch算法用于解决HMM训练的问题，即HMM参数重估的问题。给定一个观察值序列$O = o_1 o_2 \dots o_T$，该算法能确定一个$\lambda = (\pi, A, B)$，使概率值$P(O|\lambda)$最大。

定义$\xi_t(i, j)$为给定训练序列O和模型λ时，HMM在时刻t处于状态θ_i，在t+1时刻处于状态θ_j的概率，即 $\xi_t(i, j) = P(S_t = \theta_i, S_{t+1} = \theta_j | O, \lambda)$

根据Forward-Backward算法，可推导出：

$$\xi_t(i, j) = \frac{\alpha_t(i) a_{ij} b_j(o_t) \beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_t(i) a_{ij} b_{ij}(o_t) \beta_{t+1}(j)} = \frac{\alpha_t(i) a_{ij} b_j(o_t) \beta_{t+1}(j)}{P(\mathrm{O}|\lambda)}$$

定义$\gamma_t(i)$为在时刻t时处于状态θ_i的概率：$\gamma_t(i) = \sum_{j=1}^{N} \xi_t(i, j)$

$$\gamma_t(i) = \sum_{j=1}^{N} \alpha_t(i) a_{ij} b_j(o_{t+1}) \beta_{t+1}(j) / P(O|\lambda) = \alpha_t(i) \beta_t(i) / P(O|\lambda)$$

模型训练问题的解法：Baum-Welch算法

从状态θ_i向其它状态转移的次数的期望值为$\sum_{t=1}^{T-1}\gamma_t(i)$，从状态$\theta_i$转移到状态$\theta_j$的次数的期望值为$\sum_{t=1}^{T-1}\xi_t(i,j)$

可导出重估公式：

$$\overline{a_{ij}} = \sum_{t=1}^{T-1}\xi_t(i,j) \Big/ \sum_{t=1}^{T-1}\gamma_t(i)$$

$$\overline{b_{ij}(k)} = \sum_{\substack{t=1 \\ o_t=k}}^{T-1}\xi_t(i,j) \Big/ \sum_{t=1}^{T-1}\xi_t(i,j)$$

模型训练问题的解法：Baum-Welch算法

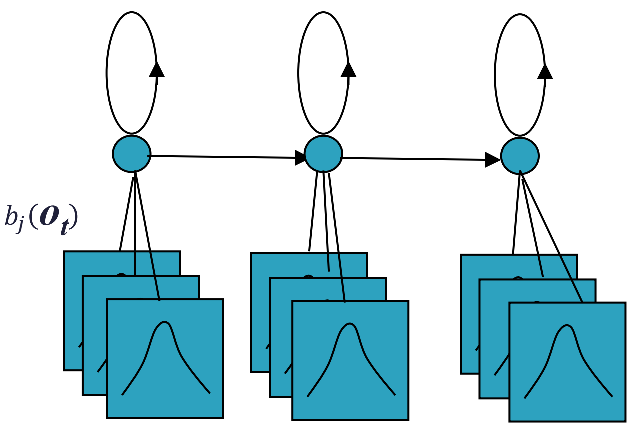
1. 参数初始化。π 和A的初值对结果影响不大，只要满足约束条件可随机选取或均值选取。B的初值对参数重估影响较大，选取算法较复杂。
2. 给定一个训练数据(观察值序列) $O = o_1 o_2 ... o_T$ ，由初始模型计算出 $\gamma_t(i)$ 和 $\xi_t(i, j)$，进一步由重估公式计算 $\overline{a_{ij}}$ 和 $\overline{b_{ij}(k)}$ 。
3. 再给定新的训练数据，把 $\overline{a_{ij}}$ 和 $\overline{b_{ij}(k)}$ 作为初始模型参数，由重估公式再次重新计算 $\overline{a_{ij}}$ 和 $\overline{b_{ij}(k)}$。
4. 重复第3步的操作，直到 $\overline{a_{ij}}$ 和 $\overline{b_{ij}(k)}$ 收敛为止，即 $P(O | \lambda)$ 趋于稳定，不再明显增大。

一般反复训练10次即可。若要训练一个优良的HMM，至少需要几十个同类的训练数据。

GMM-HMM

- 高斯混合概率密度函数（pdf）：
  - 每个状态j表示为若干函数$N(o_t)$的线性组合
  - $N(o_t)$是连续高斯概率密度函数

$$b_j(o_t) = \sum_{k=1}^{K} c_{jk} N(o_t, \mu_{jk} \Sigma_{jk})$$



GMM-HMM参数重估

- 需要重估的参数：
  - 各状态中不同pdf的权
  - 各状态中不同pdf的均值和方差
  - 转移概率

GMM-HMM参数重估

若观察概率为混合高斯分布形式，定义

$$\gamma_t^c(j,k) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right]\left[\frac{c_{jk}N(\mathbf{o}_t^c,\boldsymbol{\mu}_{jk},\boldsymbol{\Sigma}_{jk})}{\sum_{k=1}^{K}c_{jk}N(\mathbf{o}_t^c,\boldsymbol{\mu}_{jk},\boldsymbol{\Sigma}_{jk})}\right]$$

$$= \begin{cases} \dfrac{1}{P(\mathbf{O}\mid\lambda)}\pi_j\beta_1(j)c_{jk}N(o_1^c,\mu_{jk},U_{jk}), & t=1 \\ \dfrac{1}{P(\mathbf{O}\mid\lambda)}\sum_{i=1}^{N}\alpha_{t-1}(i)a_{ij}\beta_t(j)c_{jk}N(o_t^c,\mu_{jk},U_{jk}), & t>1. \end{cases}$$

GMM-HMM参数重估

则重估公式写为

$$c_{jk}=\frac{\sum_{c=1}^{C}\sum_{t=1}^{T_c}\gamma_t^c(j,k)}{\sum_{k=1}^{K}\sum_{c=1}^{C}\sum_{t=1}^{T_c}\gamma_t^c(j,k)}$$

$$\boldsymbol{\mu}_{jk}=\frac{\sum_{c=1}^{C}\sum_{t=1}^{T_c}\gamma_t^c(j,k)\mathbf{o}_t^c}{\sum_{c=1}^{C}\sum_{\mathrm{t}=1}^{T_c}\gamma_t^c(j,k)}$$

其中C为训练条数

$$\boldsymbol{\Sigma}_{jk}=\frac{\sum_{c=1}^{C}\sum_{t=1}^{T_c}\gamma_t^c(j,k)(\mathbf{o}_t^c-\boldsymbol{\mu}_{jk})(\mathbf{o}_t^c-\boldsymbol{\mu}_{jk})'}{\sum_{c=1}^{C}\sum_{t=1}^{T_c}\gamma_t^c(j,k)}$$

转移概率参数重估

$$a_{ij} = \frac{\sum_{c=1}^{C} \frac{1}{P_c} \sum_{t=1}^{T_c-1} \alpha_t^c(i) a_{ij} b_j(\mathbf{o}_{t+1}^{(c)}) \beta_{t+1}^c(j)}{\sum_{c=1}^{C} \frac{1}{P_c} \sum_{t=1}^{T_c-1} \alpha_t^c(i) \beta_t^c(i)} = \frac{\sum_{c=1}^{C} \sum_{t=1}^{T_c-1} \xi_t^c(i,j)}{\sum_{c=1}^{C} \sum_{t=1}^{T_c-1} \gamma_t^c(i)}$$

Thank you!

Any questions?