第4章 隐马尔可夫模型

宗 成 庆

中国科学院自动化研究所

cqzong@nlpr.ia.ac.cn

本章内容

1. 马尔科夫模型
2. 隐马尔可夫模型
3. 隐马模型应用
4. 习题

1. 马尔科夫模型

◆马尔可夫(Andrei Andreyevich Markov)

前苏联数学家，**切比雪夫**(1821年5月16日~1894年12月8日)的学生。在概率论、数论、函数逼近论和微分方程等方面卓有成就。他提出了用数学分析方法研究自然过程的一般图式—马尔可夫链，并开创了随机过程(马尔可夫过程)的研究。

(1856.6.14 ~ 1922.7.20)

1. 马尔科夫模型

◆马尔可夫模型描述

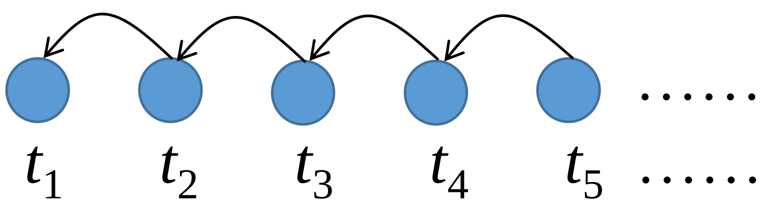
存在一类重要的随机过程：如果一个系统有N个状态S_1, $S_2,\ldots,S_N$，随着时间的推移该系统从某一状态转移到另一状态。如果用 q_t 表示系统在时间 t 的状态变量，那么，t 时刻的状态取值为 S_j ($1\leq j\leq N$) 的概率取决于前t-1个时刻 (1, 2, …, t-1)的状态，该概率为：

$$p(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \cdots)$$

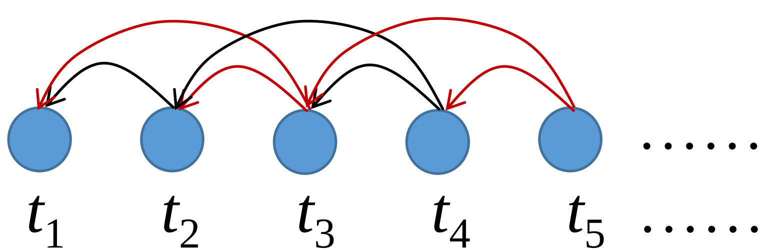
1. 马尔科夫模型

- **假设1：** 在特定情况下，t时刻的状态只与其在t-1时刻的状态相关，则该系统构成一个离散的一阶马尔可夫链(First-order Markov chain)：

$$p(q_t = S_j \mid q_{t-1} = S_i, q_{t-2} = S_k, \cdots) = p(q_t = S_j \mid q_{t-1} = S_i) \qquad \cdots (1)$$



以此类推，如果t 时刻的状态与其在t-1 和 t-2 时刻的状态相关，则该系统构成二阶马尔可夫链(Second-order Markov chain)，等等。



三阶、四阶……

1. 马尔科夫模型

- **假设2：** 如果只考虑公式(1)独立于时间 t 的随机过程，即所谓的不动性假设，状态与时间无关，那么：

$$p(q_t = S_j \mid q_{t-1} = S_i) = a_{ij}, \qquad 1 \le i, j \le N \qquad \ldots (2)$$

状态转移概率 a_{ij} 必须满足下列基本约束：

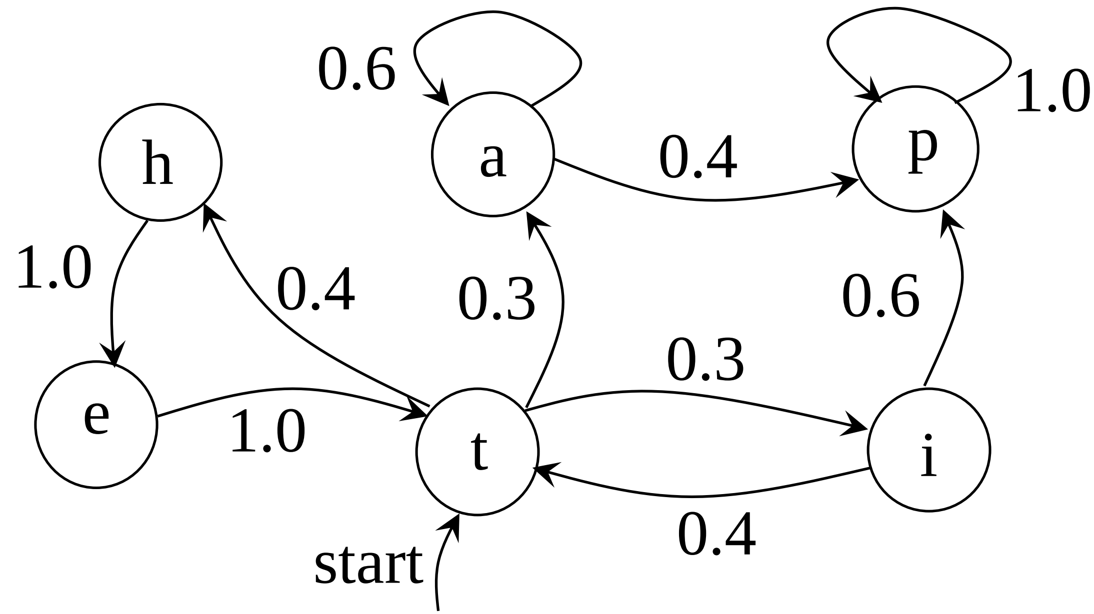
$$\begin{cases} a_{ij} \ge 0 & \ldots (3) \\ \sum_{j=1}^{N} a_{ij} = 1 & \ldots (4) \end{cases}$$

该类随机过程称为**马尔可夫模型**(Markov Model)。

1. 马尔科夫模型

◆回顾NFA

马尔可夫模型又可视为**随机的**有限状态自动机，该有限状态自动机的每一个状态转换过程都有一个相应的概率，该概率表示自动机采用这一状态转换的可能性。



- 零概率的转移弧省略。
- 每个节点上所有发出弧的概率之和等于1。

1. 马尔科夫模型

如何计算某一个状态序列 $S_1, \ldots, S_T$ 的概率？

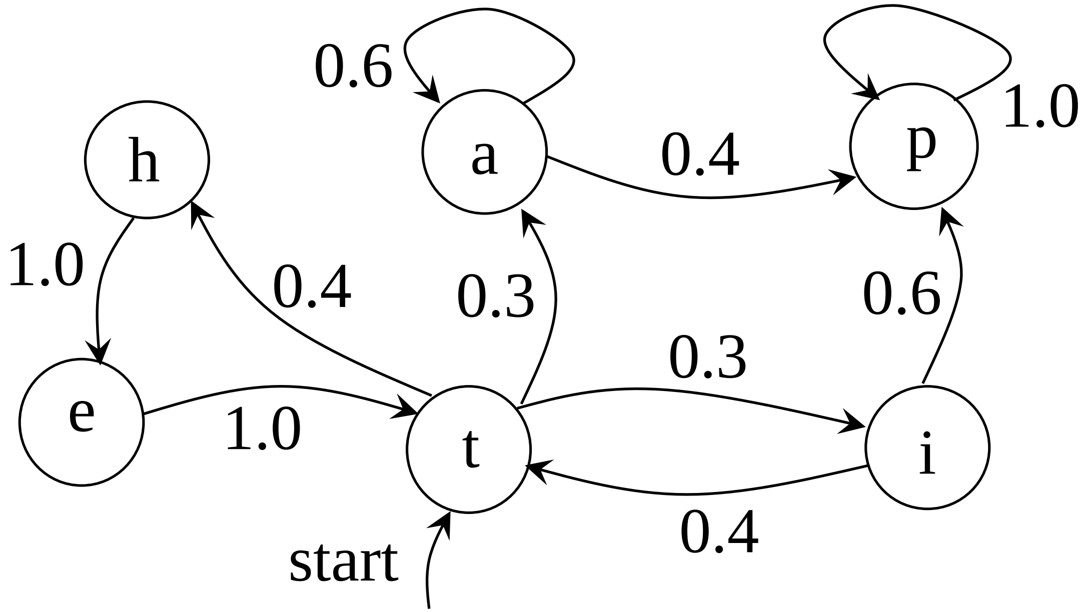
$$p(S_1,\cdots,S_T) = p(S_1)\times p(S_2 \mid S_1)\times p(S_3 \mid S_1,S_2)\times\cdots\times p(S_T \mid S_1,\cdots,S_{T-1})$$

$$= p(S_1)\times p(S_2 \mid S_1)\times p(S_3 \mid S_2)\times\cdots\times p(S_T \mid S_{T-1})$$

$$= \pi_{S_1}\prod_{t=1}^{T-1} a_{S_t S_{t+1}} \qquad \ldots (5)$$

其中，$\pi_{Si}=p(q_1=S_i)$，为初始状态的概率。

1. 马尔科夫模型



$p(\mathrm{t, i, p}) = ?$

$$= p(S_1 = \mathrm{t}) \times p(S_2 = \mathrm{i} \mid S_1 = \mathrm{t}) \times p(S_3 = \mathrm{p} \mid S_2 = \mathrm{i})$$

$$= 1.0 \times 0.3 \times 0.6$$

$$= 0.18$$

本章内容

1. 马尔科夫模型
2. 隐马尔可夫模型
3. 隐马模型应用
4. 习题

2. 隐马尔科夫模型

◆隐马尔可夫模型(Hidden Markov Model, HMM)

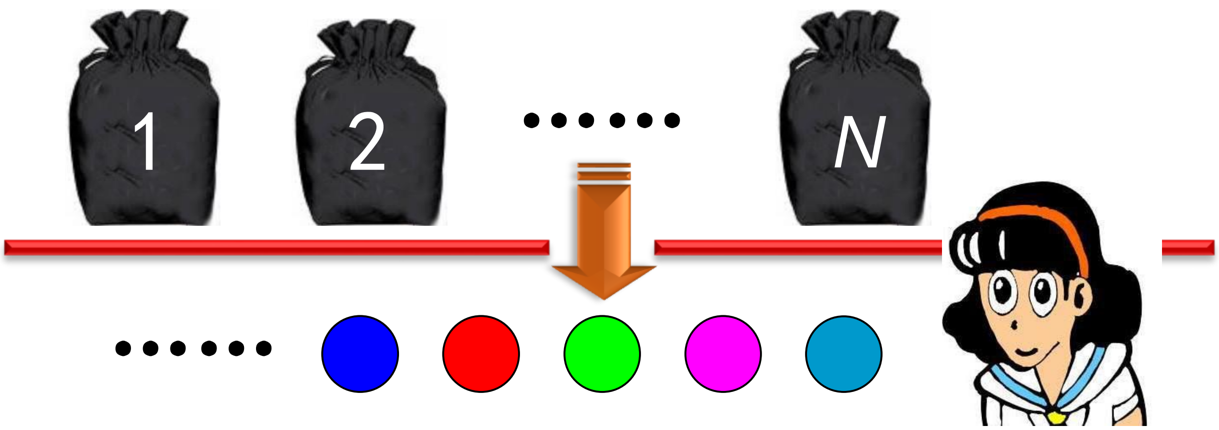
是20世纪70年代美国数学家鲍姆(Leonard E. Baum)等人提出来的。

● **模型描写**

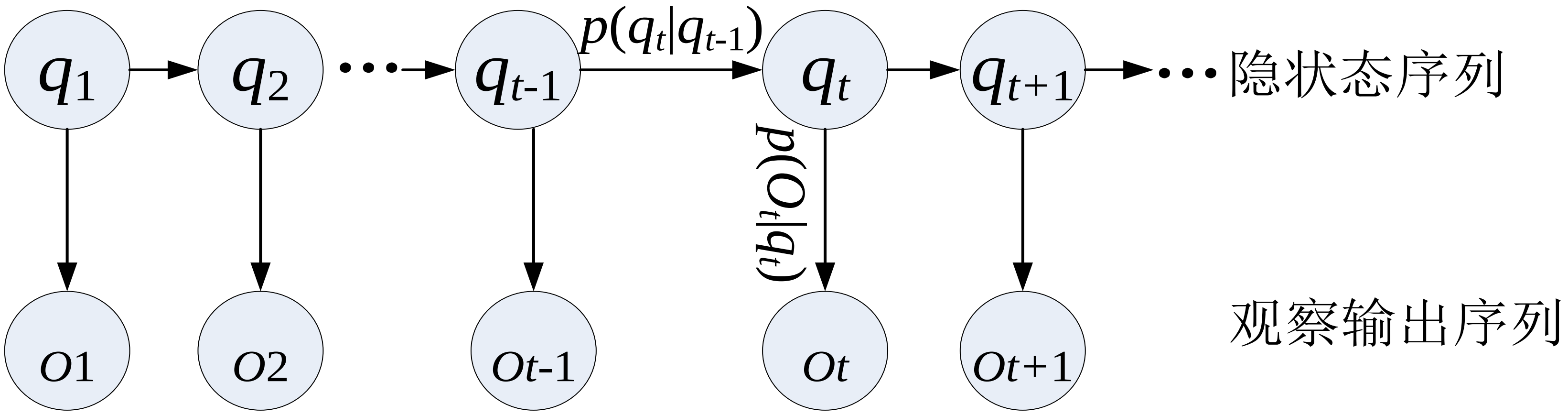
该模型描述的是一个双重随机过程，我们不知道具体的状态序列，只知道状态转移的概率，即模型的状态转换过程是不可观察的（隐蔽的），而可观察事件的随机过程是隐蔽状态转换过程的随机函数。

2. 隐马尔科夫模型

<u>例如</u>：N 个袋子，每个袋子中有 M 种不同颜色的球。一实验员根据某一概率分布选择一个袋子，然后根据袋子中不同颜色球的概率分布随机取出一个球，并报告该球的颜色。对局外人：可观察的过程是不同颜色球的序列，而袋子的序列是不可观察的。每只袋子对应HMM中的一个状态；球的颜色对应于 HMM 中状态的输出。



2. 隐马尔科夫模型



- **HMM 的组成**

  (1) 模型中的状态数为 N (袋子的数量)

  (2) 从每个状态可输出的不同符号数 M (不同颜色球的数目)

2. 隐马尔科夫模型

(3)状态转移概率矩阵$A=a_{ij}$，a_{ij}为实验员从一只袋子(状态S_i)转向另一只袋子(状态 S_j)取球的概率。其中,

$$\begin{cases} a_{ij} = p(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \le i, j \le N \\ a_{ij} \ge 0 \\ \sum_{j=1}^{N} a_{ij} = 1 \end{cases} \qquad \dots (6)$$

(4)从状态S_j观察到某一特定符号(输出)v_k 的概率分布矩阵为：$B=b_j(k)$, 其中, $b_j(k)$为从第j个袋子中取出第k种颜色球的概率。那么，

$$\begin{cases} b_j(k) = p(O_t = v_k \mid q_t = S_j), \quad 1 \le j \le N, \quad 1 \le k \le M \\ b_j(k) \ge 0 \\ \sum_{k=1}^{M} b_j(k) = 1 \end{cases} \qquad \dots (7)$$

2. 隐马尔科夫模型

(5) 初始状态的概率分布为：$\pi = \pi_i$，其中，

$$\begin{cases} \pi_i = p(q_1 = S_i), \quad 1 \le i \le N \\ \pi_i \ge 0 \\ \sum_{i=1}^{N} \pi_i = 1 \end{cases} \qquad \dots (8)$$

为了方便起见，一般将 HMM记为：$\mu = (A, B, \pi)$，或者 $\mu = (S, O, A, B, \pi)$，用以指出模型的参数集合。

2. 隐马尔科夫模型

- **问题**

给定模型$\mu=(A, B, \pi)$，如何产生观察序列 $O=O_1O_2\cdots O_T$ 呢？

(a) 令 $t=1$;

(b) 根据初始状态分布π_i 选择初始状态 $q_1=S_i$;

(c) 根据状态 S_i 的输出概率分布 $b_i(k)$, 输出 $O_t=v_k$;

(d) 根据状态转移概率a_{ij}，转移到新状态 $q_{t+1}=S_j$;

(e) $t=t+1$, 如果 $t<T$, 重复步骤 (c) (d), 否则结束。

2. 隐马尔科夫模型

- **三个问题**

(1) 在给定模型μ=(A, B, π) 和观察序列 $O＝O_1O_2 \ldots O_T$ 的情况下，怎样快速计算概率 $p(O|\mu)$?

(2) 在给定模型 μ=(A, B, π) 和观察序列 $O＝O_1O_2 \ldots O_T$的情况下，如何选择在一定意义下“最优”的状态序列 $Q = q_1 q_2 \ldots q_T$，使该状态序列“最好地解释”观察序列?

(3) 给定一个观察序列 $O＝O_1O_2 \ldots O_T$，如何根据最大似然估计求模型的参数值？或者说如何调节模型的参数，使得 $p(O|\mu)$ 最大?

2. 隐马尔科夫模型

◆问题求解

● **求解问题1**：快速计算观察序列概率$p(O|\mu)$

对于给定的模型 $\mu=(A, B, \pi)$，$p(O|\mu)=?$

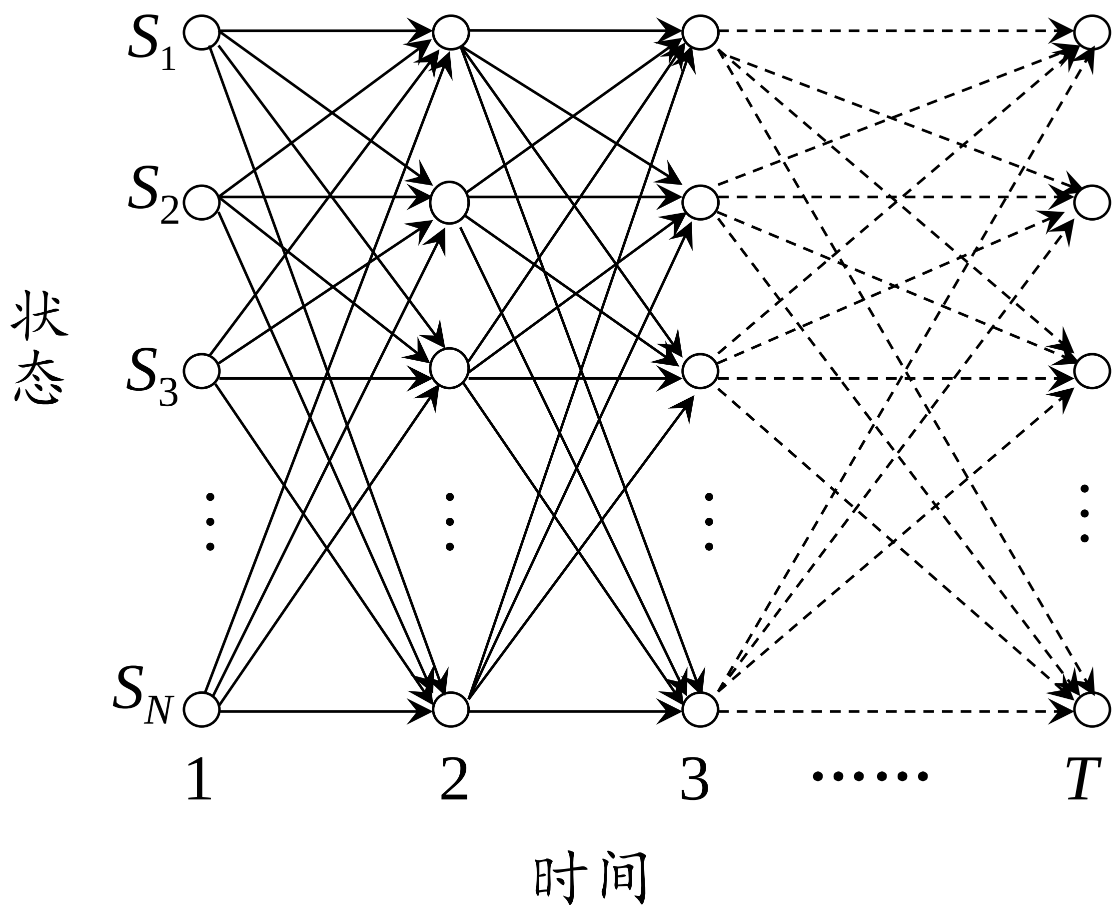
$$p(O|\mu)=\sum_{Q} p(O,Q|\mu)$$

$$=\sum_{Q} \boxed{p(Q|\mu)} \times \boxed{p(O|Q,\mu)} \qquad \ldots (9)$$

$$p(Q|\mu)=\pi_{q_1} \times a_{q_1 q_2} \times a_{q_2 q_3} \times \cdots \times a_{q_{t-1} q_T} \qquad \ldots (10)$$

$$p(O|Q,\mu)=b_{q_1}(O_1) \times b_{q_2}(O_2) \times \cdots \times b_{q_T}(O_T) \qquad \ldots (11)$$

2. 隐马尔科夫模型



困难：

如果模型μ有N个不同的状态，时间长度为T，那么有N^T个可能的状态序列，搜索路径成指数级组合爆炸。

2. 隐马尔科夫模型

- **解决思路：** 采用“化整为零”，动态规划的求解策略。

方法①：定义**前向变量**$\alpha_t(i)$，从1时刻开始，依次计算到达时刻 t 时形成的输出序列的概率：

$$\alpha_t(i) = p(O_1 O_2 \cdots O_t, q_t = S_i \mid \mu) \qquad \ldots(12)$$

如果可以计算$\alpha_t(i)$，就可以高效地求得 $p(O|\mu)$。

2. 隐马尔科夫模型

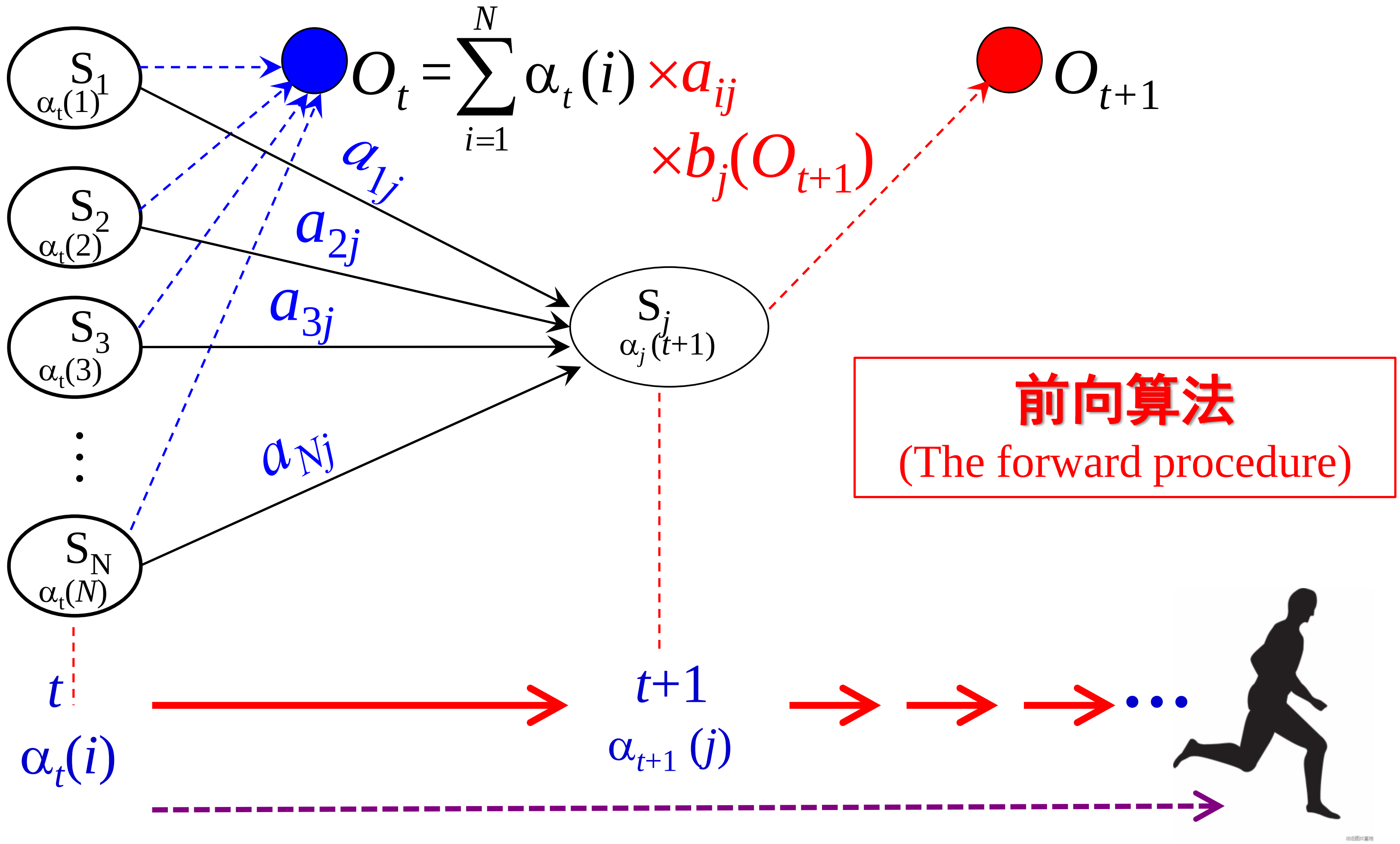
由于所有状态产生的输出都有可能成为该观察序列中的一元，而我们需要计算到达状态 q_T 时观察到序列$O=O_1O_2\cdots O_T$ 的概率，所以，

$$p(O\,|\,\mu)=\sum_{S_i} p(O_1O_2\cdots O_T, q_T=S_i\,|\,\mu)=\sum_{i=1}^{N}\alpha_T(i) \qquad \ldots (13)$$

在时间 t+1 的前向变量可以根据时间 t 的前向变量$\alpha_t(1), \ldots, \alpha_t(N)$ 的值递推计算：

$$\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)\times a_{ij}]\times b_j(O_{t+1}) \qquad \ldots (14)$$

2. 隐马尔科夫模型


S_1 $\alpha_t(1)$
S_2 $\alpha_t(2)$
S_3 $\alpha_t(3)$
S_N $\alpha_t(N)$
a_{1j}
a_{2j}
a_{3j}
a_{Nj}
S_j $\alpha_j(t+1)$
$O_t = \sum_{i=1}^{N} \alpha_t(i) \times a_{ij} \times b_j(O_{t+1})$
O_{t+1}
t
$\alpha_t(i)$
$t+1$
$\alpha_{t+1}(j)$
…
前向算法
(The forward procedure)

2. 隐马尔科夫模型

- **算法1：** 前向算法描述

(1) 初始化： $\alpha_1(i)=\pi_i b_i(O_1)$， $1 \le i \le N$

(2) 循环计算：

$$\alpha_{t+1}(j)=[\sum_{i=1}^{N}\alpha_t(i)a_{ij}]\times b_j(O_{t+1}),\quad 1\le t\le T-1$$

(3) 结束，输出：

$$p(O\,|\,\mu)=\sum_{i=1}^{N}\alpha_T(i)$$

2. 隐马尔科夫模型

- **算法的时间复杂性：**

每计算一个$\alpha_t(i)$ 必须考虑从 t-1 时的所有N 个状态转移到状态 S_i 的可能性，时间复杂性为 $O(N)$；每个时刻 t 要计算N 个前向变量：$\alpha_t(1)$, $\alpha_t(2)$,…,$\alpha_t(N)$，所以，时间复杂性为：$O(N)\times N = O(N^2)$。又因 t =1, 2, …, T，所以前向算法总的复杂性为：$O(N^2T)$。

2. 隐马尔科夫模型

方法②：定义**后向变量**$\beta_t(i)$，计算在给定模型$\mu=(A, B, \pi)$和假定在时间t状态为S_i的条件下，模型输出观察序列$O_{t+1}O_{t+2}\cdots O_T$的概率：

$$\beta_t(i) = p(O_{t+1}O_{t+2}\cdots O_T \mid q_t = S_i, \mu) \qquad \ldots (15)$$

2. 隐马尔科夫模型

第1步，计算从时刻 t 到 t+1，模型由状态 S_i 转移到状态 S_j，并从 S_j 输出 O_{t+1}概率：$a_{ij}\times b_j(O_{t+1})$；

第2步，计算在时刻 t+1、状态为 S_j 的条件下，模型输出观察序列 $O_{t+1}O_{t+2}\ldots O_T$的概率按后向变量的定义为：$\beta_{t+1}(j)$。

于是，有归纳关系：

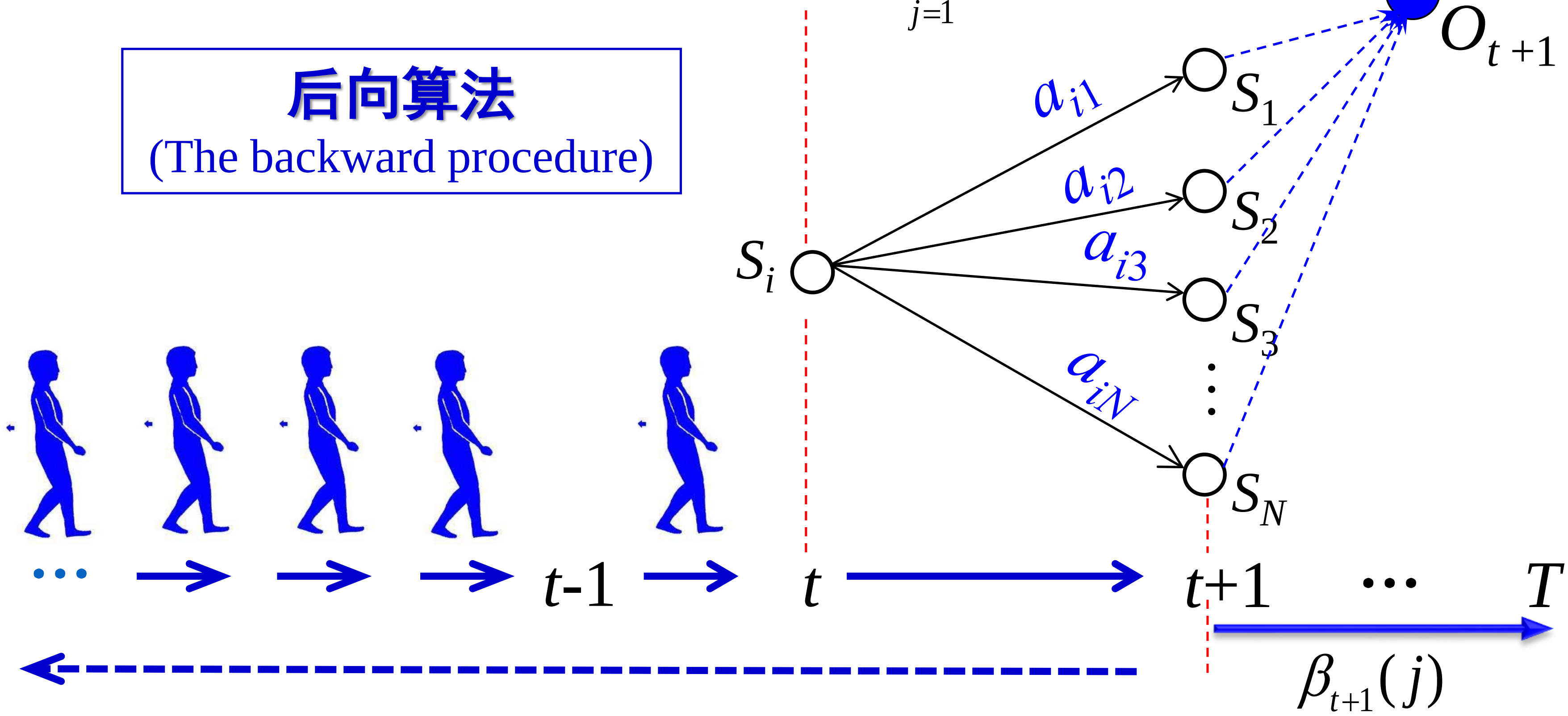
$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \times \beta_{t+1}(j) \qquad \ldots (16)$$

归纳顺序：$\beta_T(\bullet), \beta_{T\text{-}1}(\bullet), \ldots, \beta_1(\bullet)$

2. 隐马尔科夫模型

后向算法
(The backward procedure)

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \times \beta_{t+1}(j)$$



2. 隐马尔科夫模型

- **算法2**：后向算法描述

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

(2) 循环计算：

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(O_{t+1}) \times \beta_{t+1}(j), \qquad T-1 \ge t \ge 1, \quad 1 \le i \le N$$

(3) 输出结果：

$$p(O \mid \mu) = \sum_{i=1}^{N} \beta_1(i) \times \pi_i \times b_i(O_1)$$

算法的时间复杂性：$O(N^2T)$

2. 隐马尔科夫模型

- **求解问题2**：如何发现“最优”状态序列能够“最好地解释”观察序列？

  如何理解“最优”的状态序列？

解释(*a*): 对于每个时刻 t ($1\leq t \leq T$), 寻找对应观察符号概率最大的状态, 即寻找使得 $\gamma_t(i)=p(q_t=S_i|O,\mu)$ 最大的 q_t。

$$\gamma_t(i)=p(q_t=S_i\,|\,O,\mu)=\frac{p(q_t=S_i,O\,|\,\mu)}{p(O\,|\,\mu)} \qquad \ldots (17)$$

2. 隐马尔科夫模型

- **分段计算：**

(1) 模型在时刻 t 到达状态 S_i, 并且输出$O=O_1O_2\ldots O_t$。根据前向变量的定义，实现这一步的概率为$\alpha_t(i)$。

(2) 从时刻 t、状态 S_i 出发，模型输出$O=O_{t+1}O_{t+2}\ldots O_T$，根据后向变量定义，实现这一步的概率为$\beta_t(i)$。

于是：

$$p(q_t = S_i, O \mid \mu) = \alpha_t(i) \times \beta_t(i) \qquad \ldots (18)$$

2. 隐马尔科夫模型

而 $p(O|\mu)$与时间 t 的状态无关，因此：

$$p(O \mid \mu) = \sum_{i=1}^{N} \alpha_t(i) \times \beta_t(i) \qquad \ldots (19)$$

将公式(19)和(18)：$p(q_t = S_i, O|\mu) = \alpha_t(i) \times \beta_t(i)$ 带入(17)式：

$$\gamma_t(i) = p(q_t = S_i \mid O, \mu) = \frac{p(q_t = S_i, O \mid \mu)}{p(O \mid \mu)} \qquad \ldots (17)$$

得到：

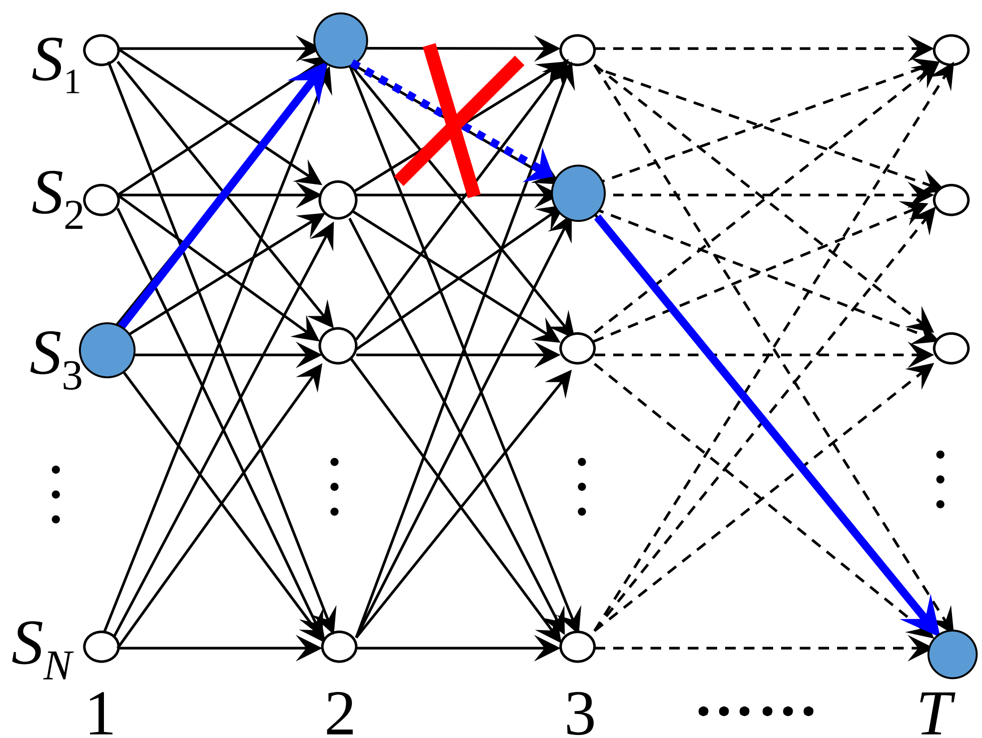
$$\gamma_t(i) = \frac{\alpha_t(i) \times \beta_t(i)}{\sum_{i=1}^{N} \alpha_t(i) \times \beta_t(i)} \qquad \ldots (20)$$

t 时刻的最优状态为：$\hat{q}_t = \arg\max_{1 \le i \le N} (\gamma_t(i))$

2. 隐马尔科夫模型

问题：每一个状态单独最优不一定使整体的状态序列最优，可能两个最优的状态 $\hat{q}_t$ 和 $\hat{q}_{t+1}$ 之间没有转移概率，即 $a_{\hat{q}_t\hat{q}_{t+1}}=0$。

结论：解释(*a*)在很多情况下不可取。



2. 隐马尔科夫模型

解释(*b*)： 在给定模型μ 和观察序列O的条件下求概率最大的状态序列：$\hat{Q} = \arg\max_{Q} p(Q \mid O, \mu)$ … (21)

采用动态规划策略, 搜索全局最优状态序列—Viterbi 搜索算法。

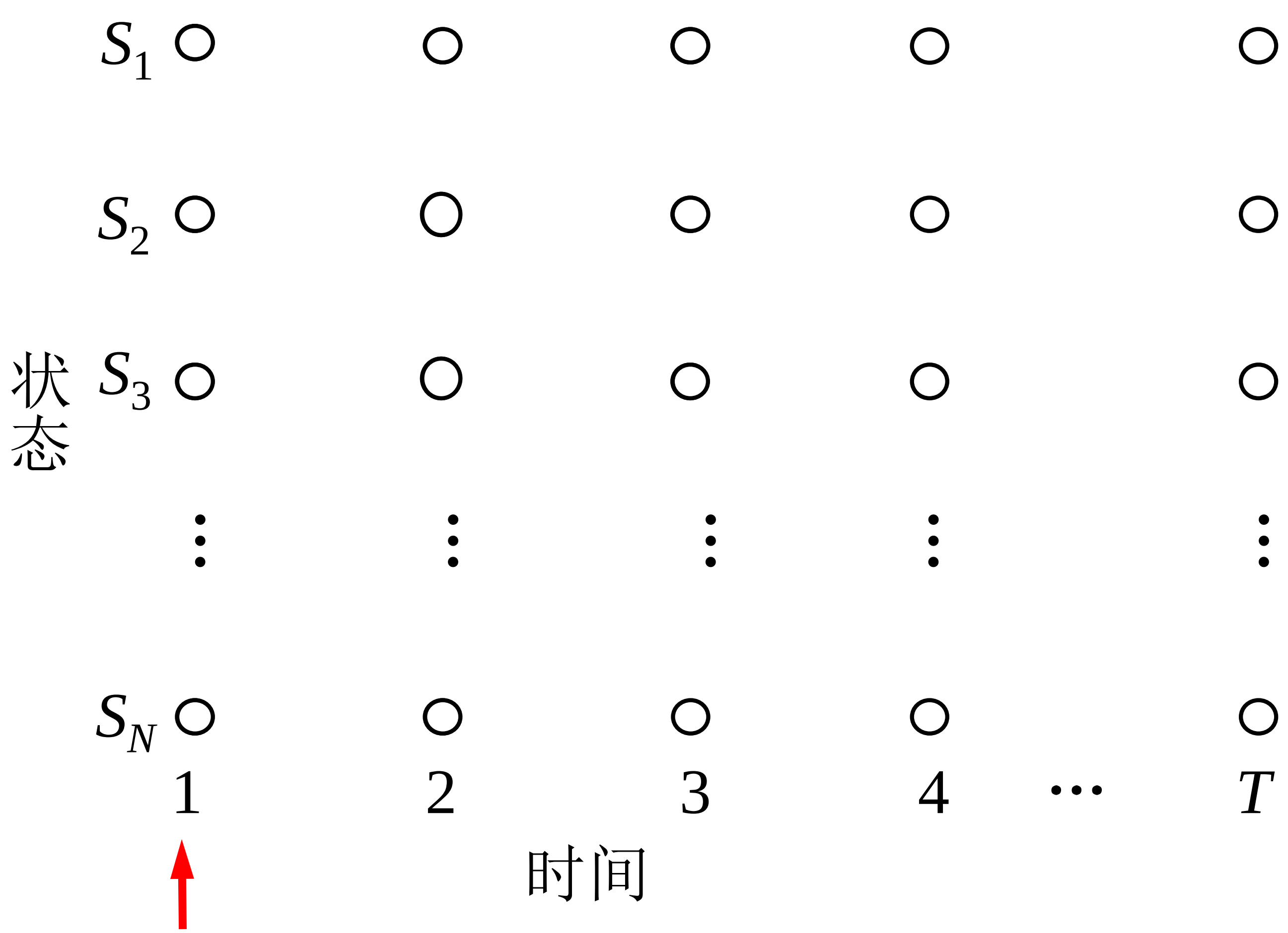
- **定义：** Viterbi变量$\delta_t(i)$是在时间t 时模型沿着某一条路径到达S_i，输出观察序列$O=O_1O_2\cdots O_t$的最大概率为：

$$\delta_t(i) = \max_{q_1, q_2, \cdots, q_{t-1}} p(q_1, q_2, \cdots, q_t = S_i, O_1 O_2 \cdots O_t \mid \mu) \quad \dots (22)$$

递归计算：$\delta_{t+1}(i) = \max_{j}[\delta_t(j) \cdot a_{ji}] \cdot b_i(O_{t+1})$ … (23)

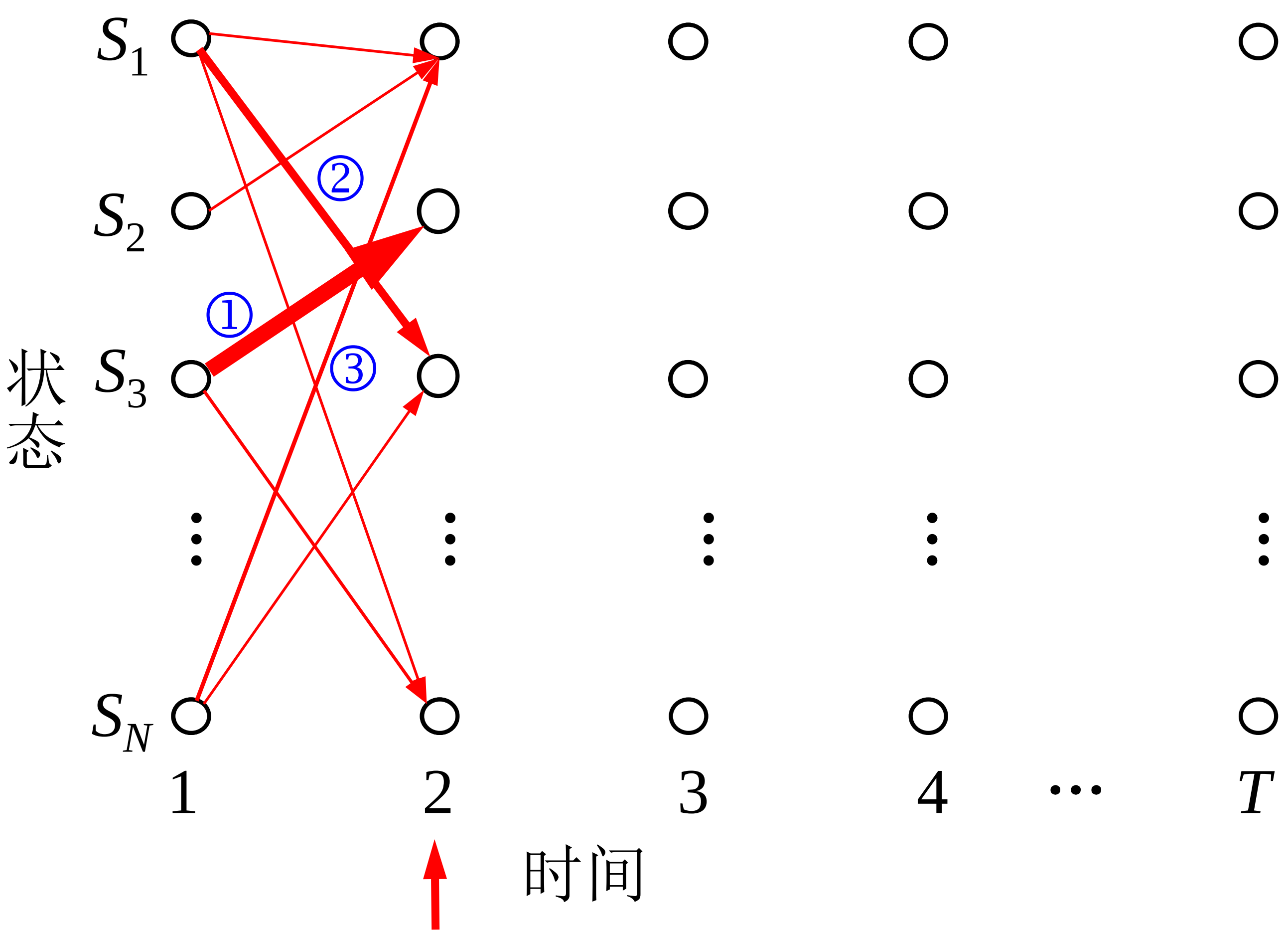
2. 隐马尔科夫模型

图解 Viterbi 搜索过程



2. 隐马尔科夫模型

图解 Viterbi 搜索过程



剪枝策略：

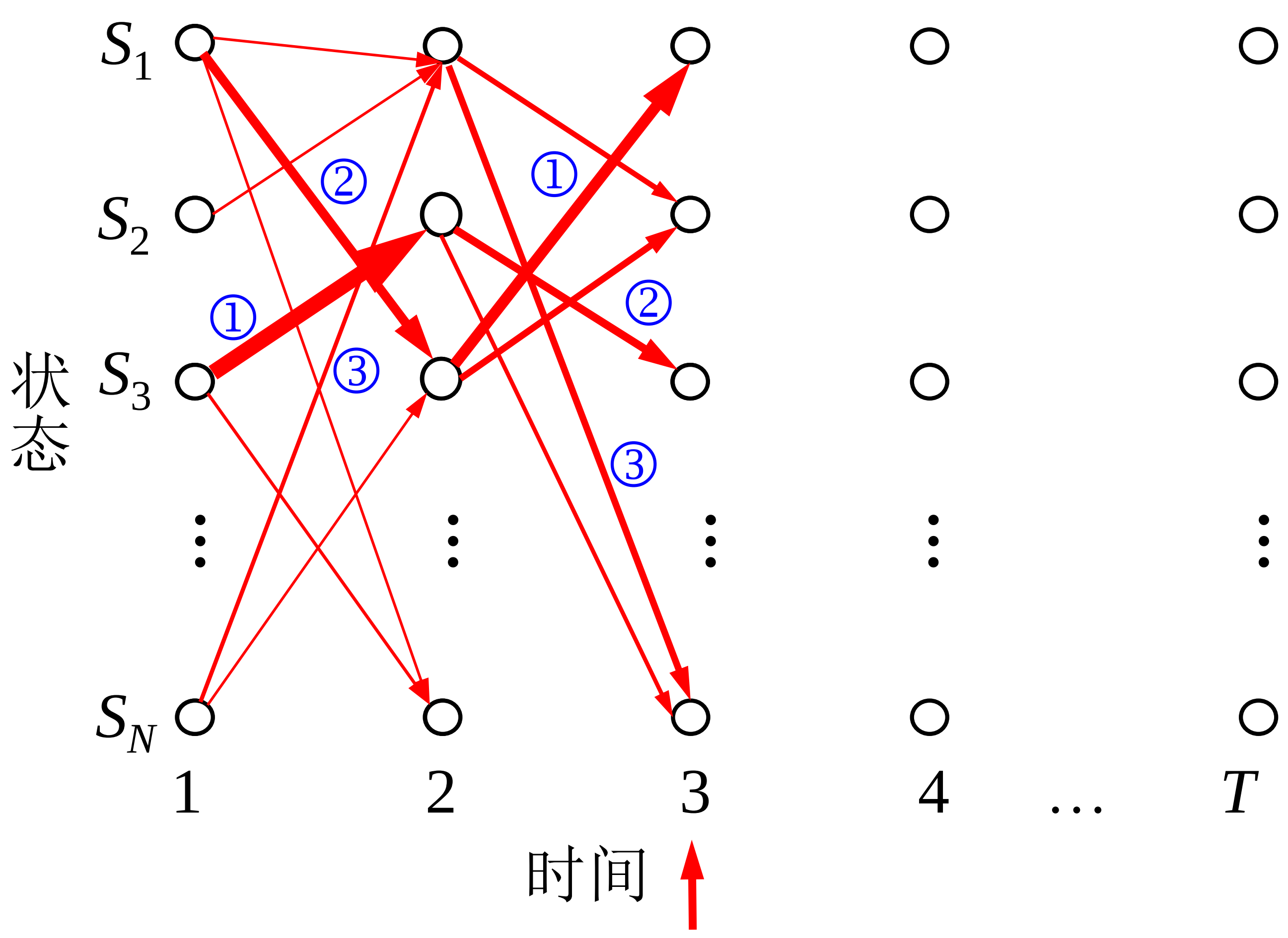
❶ $\delta_t(j) \geq \Delta$

❷ NPath $\leq \sigma$

(3)

2. 隐马尔科夫模型

图解
Viterbi
搜索
过程



剪枝策略：

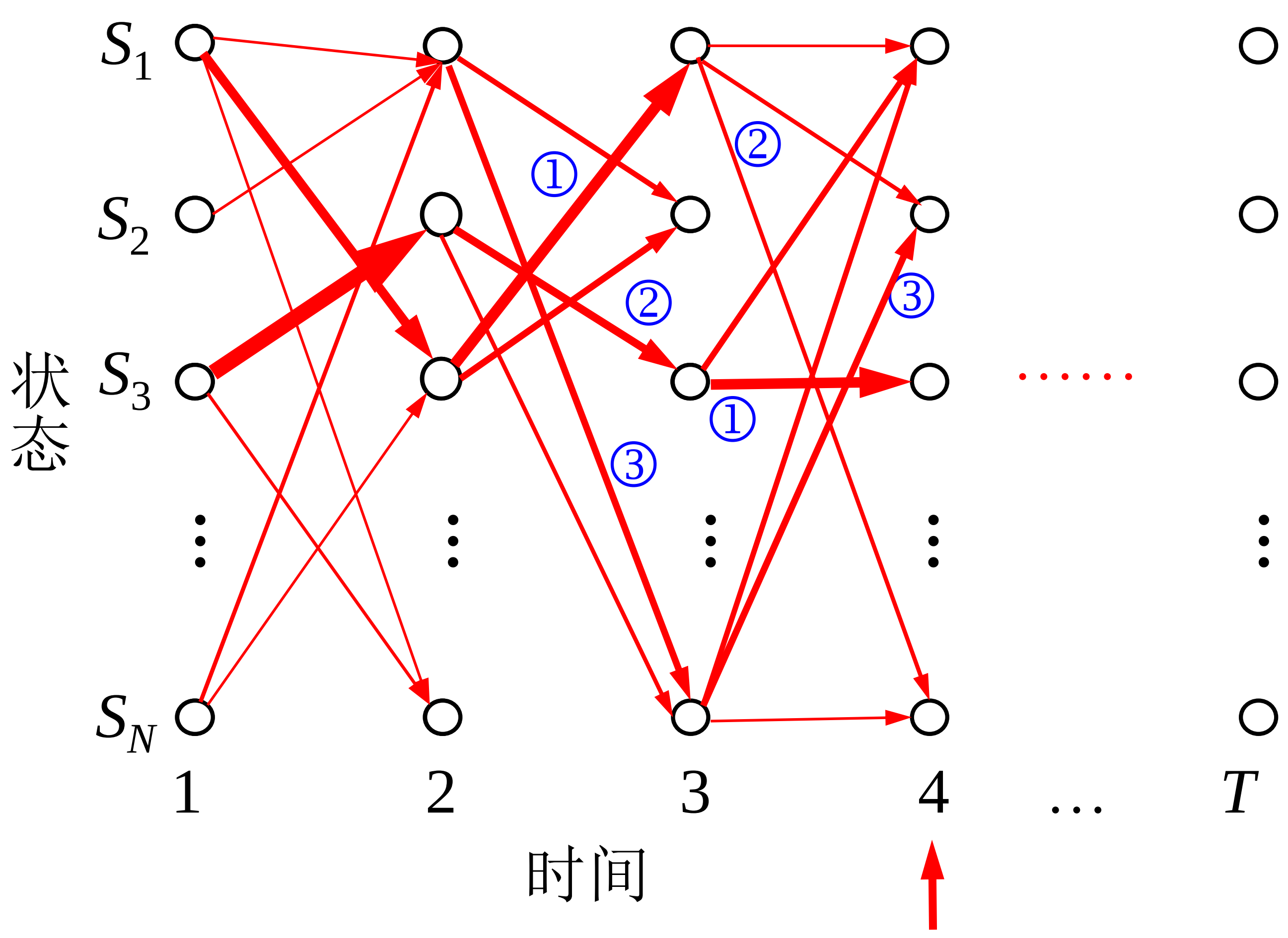
❶ $\delta_t(j) \geq \Delta$

❷ NPath $\leq \sigma$

(3)

2. 隐马尔科夫模型

图解
Viterbi
搜索
过程



剪枝策略：

❶ $\delta_t(j) \geq \Delta$

❷ NPath $\leq \sigma$

(3)

2. 隐马尔科夫模型

- **算法3：** Viterbi 算法描述

(1)初始化：$\delta_1(i) = \pi_i b_i(O_1)$， $1 \le i \le N$

概率最大的路径变量：$\psi_1(i) = 0$

(2)递推计算：

$$\delta_t(j) = \max_{1 \le i \le N}[\delta_{t-1}(i) \cdot a_{ij}] \cdot b_j(O_t), \quad 2 \le t \le T, \quad 1 \le j \le N$$

$$\psi_t(j) = \arg\max_{1 \le i \le N}[\delta_{t-1}(i) \cdot a_{ij}] \cdot b_j(O_t),\ 2 \le t \le T,\ 1 \le i \le N$$

(3)结束：$\hat{Q}_T = \arg\max_{1 \le i \le N}[\delta_T(i)],\ \hat{p}(\hat{Q}_T) = \max_{1 \le i \le N} \delta_T(i)$

(4)通过回溯得到路径（状态序列）：

$$\hat{q}_t = \psi_{t+1}(\hat{q}_{t+1}), \quad t = T-1,\ T-2,\ \cdots,\ 1$$

算法的时间复杂度：$O(N^2T)$

2. 隐马尔科夫模型

- **求解问题3：** 模型参数学习

  如何估计模型的参数 π_i, a_{ij}, $b_j(k)$, 使得观察序列 $O=O_1O_2\cdots O_T$ 的概率 $p(O|\mu)$ 最大。

前向后向算法

(Baum-Welch or forward-backward procedure)

2. 隐马尔科夫模型

➢ **情况1**：存在大量的标注样本, 观察序列O 的状态 $Q = q_1 q_2 \dots q_T$ 是已知的, 可用最大似然估计方法计算μ 的参数: $\bar{\pi}_i = \delta(q_1, S_i)$

$$\bar{a}_{ij} = \frac{Q\text{中从状态}q_i\text{转移到}q_j\text{的次数}}{Q\text{中所有从状态}q_i\text{转移到另一状态（包括}q_j\text{自身）的总数}}$$

$$= \frac{\sum_{t=1}^{T-1} \delta(q_t, S_i) \times \delta(q_{t+1}, S_j)}{\sum_{t=1}^{T-1} \delta(q_t, S_i)} \qquad \dots (24)$$

其中， $\delta(x, y)$ 为克罗奈克(Kronecker)函数， 当 $x = y$ 时, $\delta(x, y)=1$, 否则$\delta(x, y) = 0$。

2. 隐马尔科夫模型

类似地，

$$\bar{b}_j(k) = \frac{Q\text{中从状态}q_j\text{输出符号}v_k\text{的次数}}{Q\text{到达}q_j\text{的总次数}}$$

$$= \frac{\sum_{t=1}^{T} \delta(q_t, S_j) \times \delta(O_t, v_k)}{\sum_{t=1}^{T} \delta(q_t, S_j)} \qquad \ldots (25)$$

其中，v_k 是模型输出符号集中的第 k 个符号。

2. 隐马尔科夫模型

➢ **情况2：**没有大量标注的样本

● **期望值最大化算法**(**E**xpectation-**M**aximization, **EM**)

<u>基本思想</u>：初始化时随机地给模型的参数赋值(遵循限制规则，如从某一状态出发的转移概率满足非负性、总和为1的约束)，得到模型μ_0，然后可以从μ_0得到从某一状态转移到另一状态的期望次数，然后以期望次数代替公式中的实际次数，得到模型参数的新估计值，由此得到新的模型μ_1，从μ_1又可得到模型中隐变量的期望值，由此重新估计模型参数。循环这一过程，参数将收敛于最大似然估计值。

2. 隐马尔科夫模型

给定模型 μ 和观察序列$O=O_1O_2\cdots O_T$，那么，在时间 t 位于状态 S_i，时间 t+1位于状态S_j的概率：

$$\xi_t(i,j)=p(q_t=S_i,q_{t+1}=S_j\,|\,O,\mu)$$

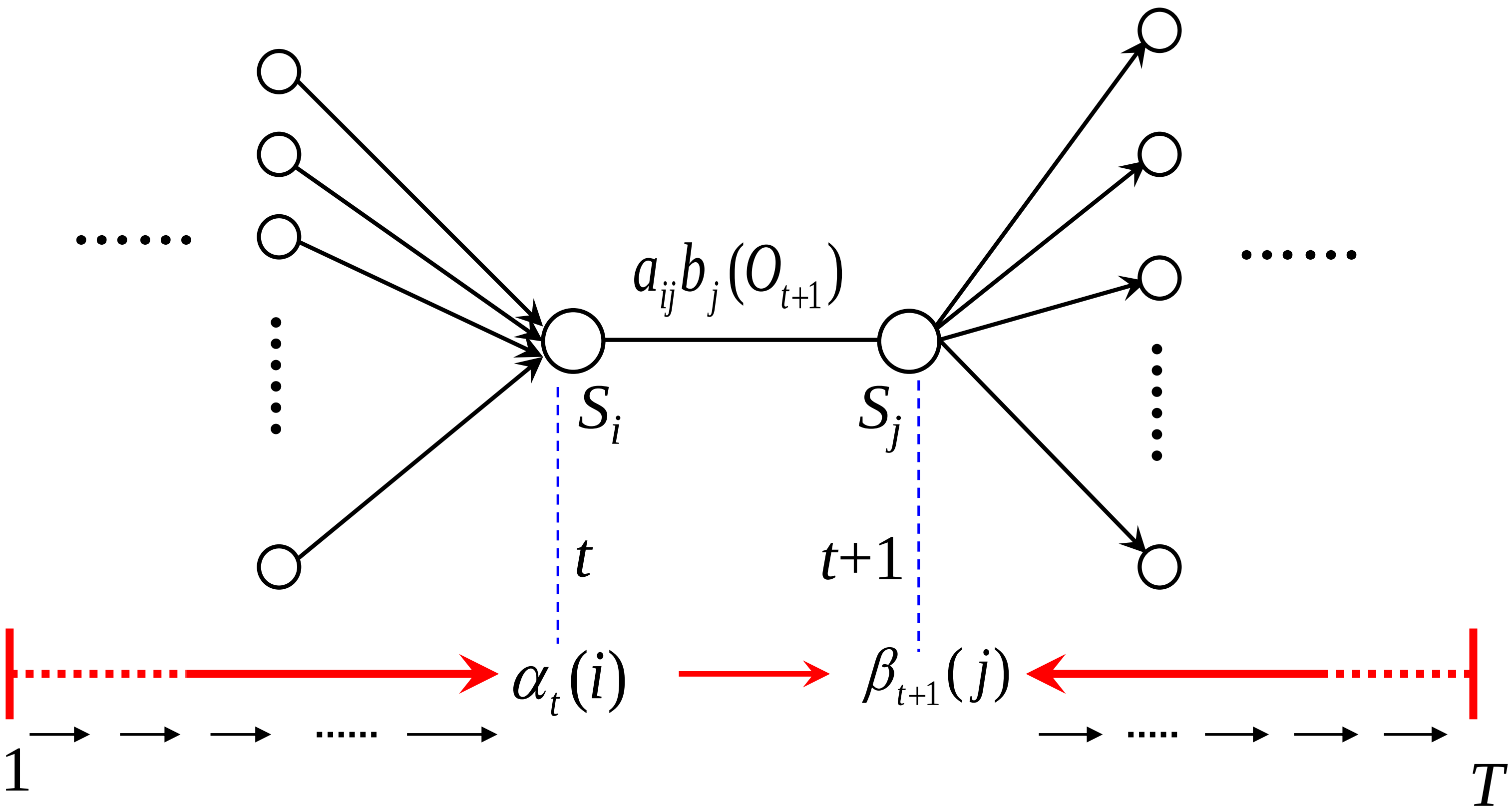
$$=\frac{p(q_t=S_i,q_{t+1}=S_j,O\,|\,\mu)}{p(O\,|\,\mu)}$$

$$=\frac{\alpha_t(i)\times a_{ij}b_j(O_{t+1})\times\beta_{t+1}(j)}{p(O\,|\,\mu)}$$

$$=\frac{\alpha_t(i)\times a_{ij}b_j(O_{t+1})\times\beta_{t+1}(j)}{\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_t(i)\times a_{ij}b_j(O_{t+1})\times\beta_{t+1}(j)} \qquad \ldots (26)$$

2. 隐马尔科夫模型

图解:



2. 隐马尔科夫模型

那么，给定模型 μ 和观察序列$O=O_1O_2\cdots O_T$，在时间 t 位于状态 S_i 的概率为：

$$\gamma_t(i)=\sum_{j=1}^{N}\xi_t(i,j) \qquad \ldots (27)$$

由此，模型 μ 的参数可由下面的公式重新估计：

(i) q_1 为 S_i 的概率：

$$\pi_i=\gamma_1(i) \qquad \ldots (28)$$

2. 隐马尔科夫模型

(ii)

$$a_{ij}=\frac{Q\text{中从状态}q_i\text{转移到}q_j\text{的期望次数}}{Q\text{中所有从状态}q_i\text{转移到下一状态(包括}q_j\text{自身)的期望次数}}$$

$$=\frac{\sum_{t=1}^{T-1}\xi_t(i,j)}{\sum_{t=1}^{T-1}\gamma_t(i)} \qquad \ldots (29)$$

(iii)

$$b_j(k)=\frac{Q\text{中从状态}q_j\text{输出符号}v_k\text{的期望次数}}{Q\text{到达}q_j\text{的期望次数}}$$

$$=\frac{\sum_{t=1}^{T}\gamma_t(j)\times\delta(O_t,v_k)}{\sum_{t=1}^{T}\gamma_t(j)} \qquad \ldots (30)$$

2. 隐马尔科夫模型

- **算法4：** Baum-Welch 算法(前向后向算法)描述:

(1)初始化： 随机给 π_i, a_{ij}, $b_j(k)$赋值，满足如下约束：

$$\begin{cases} \sum_{i=1}^{N} \pi_i = 1 & \\ \sum_{j=1}^{N} a_{ij} = 1 & 1 \le i \le N \\ \sum_{k=1}^{M} b_i(k) = 1 & 1 \le i \le N \end{cases} \quad \ldots (31)$$

由此得到模型 μ_0，令 $i = 0$。

2. 隐马尔科夫模型

(2)执行 EM 算法:

E-步： 由模型μ_i根据公式(26)和(27)计算期望值$\xi_t(i,j)$和$\gamma_t(i)$。

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

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

M-步： 用E-步中所得到的期望值，根据公式 (28~30) 重新估计 π_i, a_{ij}, $b_j(k)$ 得到模型μ_{i+1}。

循环： $i=i+1$，重复执行 E-步和M-步，直至π_i, a_{ij}, $b_j(k)$的值收敛：$|\log p(O|\mu_{i+1})-\log p(O|\mu_i)|<\varepsilon$。

(3)结束算法，获得相应的参数。

2. 隐马尔科夫模型

- **注意：**
  - Viterbi 算法运算中的小数连乘和 Baum-Welch 算法的小数运算出现溢出现象，通常取对数或者乘以放大系数。

- **参考：**
  - HTK 开源代码：http://htk.eng.cam.ac.uk/
  - Rabiner, L. R. and B. H. Juang. 1993. ***Fundamentals of Speech Recognition***. Prentice-Hall. Pages 365-368

本章内容

1. 马尔科夫模型
2. 隐马尔可夫模型
3. 隐马模型应用
4. 习题

3. 隐马模型应用

- 出生于捷克一个富有的犹太家庭
- 1949 年移民美国
- 麻省理工学院获博士学位，之后在哈佛大学任教，一年后到康乃尔大学任教
- 1972年到IBM 华生实验室(IBM T.G.Watson Labs)做学术休假，领导了语音识别实验室，两年后离开康奈尔大学正式去IBM工作
- 提出了**基于因马尔科夫模型的语音识别方法，建立了*n*-gram 语言模型**，为此当选美国工程院院士
- 1990S离开IBM公司，到约翰霍普金斯大学任教，建立了世界著名的 CLSP 实验室。

——选自百度百科(https://baike.baidu.com/item/贾里尼克)

弗莱德里克.贾里尼克
(Frederek Jelinek)
(1932.11.18 ~ 2010.9.14)

3. 隐马模型应用

◆以汉语自动分词（实体识别）和词性标注为例。

例如：武汉市长江大桥于1957年9月6日竣工。

列出所有可能的切分和词性标注结果：

①武汉市/N 长江/N 大桥/N 于/P 1957年/Dat 9月/Dat 6日/Dat 竣工/V。/Pun

②武汉/N 市长/N 江大桥/N 于/P 1957年/Dat 9月/Dat 6日/Dat 竣工/V。/Pun

N_f, 姓氏

如何用 HMM 解决问题这一问题？

3. 隐马模型应用

(1) 如何确定状态及其数目？

(2) 如何观察及其各自的数目？

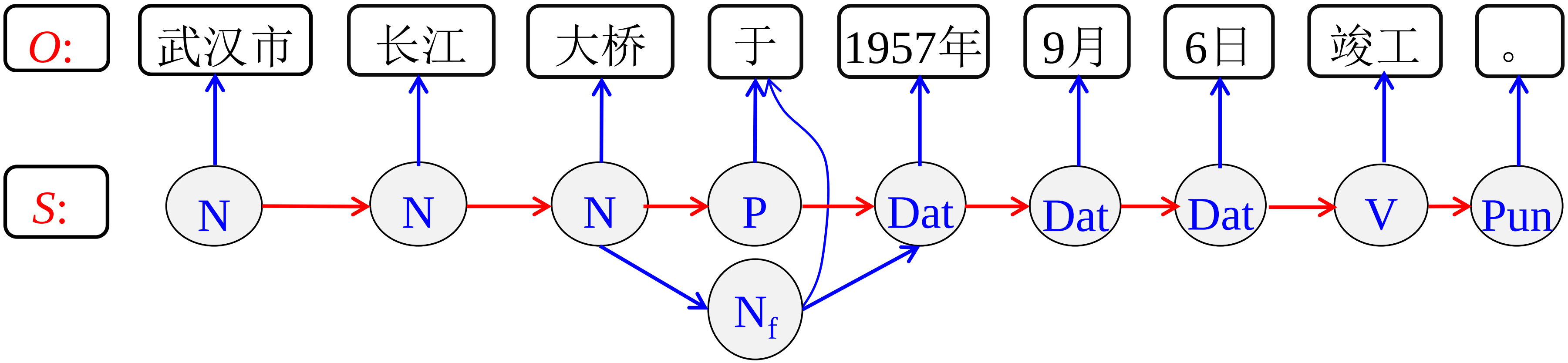
(3) 如何估计参数：初始状态概率、状态转移概率、输出概率？

- **思路:**

　　如果把汉语自动分词结果作为观察序列 $O=O_1O_2\ldots O_T$，那么，对于分词而言，我们需要求解：$\hat{O} = \arg\max\limits_{O} p(O\,|\,\mu)$，而对于词性标注而言，则需求解：$\hat{Q} = \arg\max\limits_{Q} p(Q\,|\,O,\mu)$。

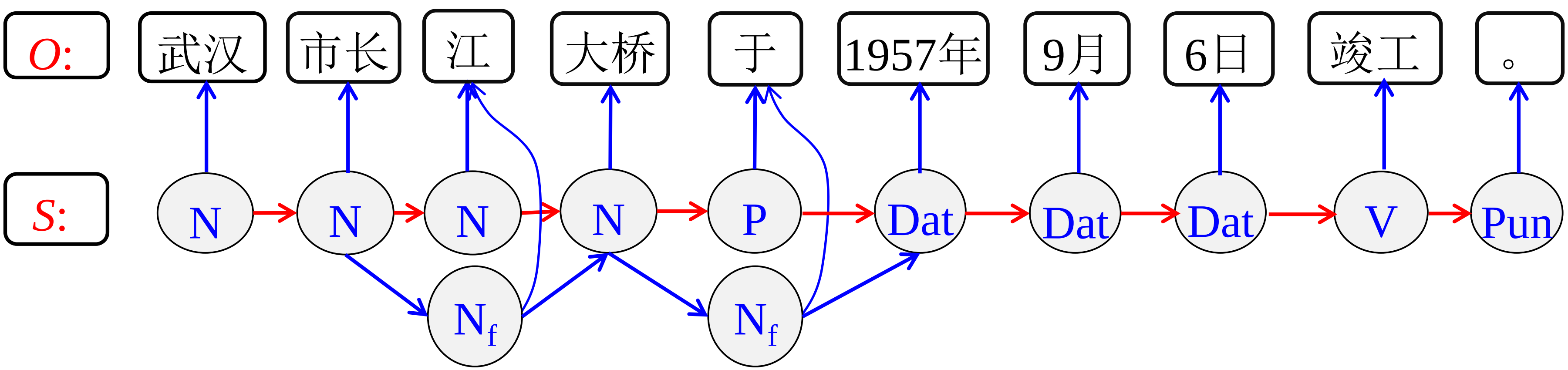
<u>进一步解释</u>：利用HMM模型μ =(A, B, π)，对于任意给定的输入句子，在所有可能的词序列O中求解使概率$p(O|\mu)$最大的候选，并快速地选择“最优”的词性序列，使其最好地解释分词结果。

3. 隐马模型应用



①武汉市/N 长江/N 大桥/N 于/P 1957年/Dat 9月/Dat 6日/Dat 竣工/V。/Pun

②武汉市/N 长江/N 大桥/N 于/N_f 1957年/Dat 9月/Dat 6日/Dat 竣工/V。/Pun



3. 隐马模型应用

● 模型参数

- **观察序列：** 单词序列
- **状态序列：** 词类标记序列
- **状态数目*N*：** 为词类标记符号的个数，如北大语料库词类标记，一级标记个数为26，三级标记数为106
- **输出符号数*M*：** 每个状态可输出的不同词汇个数，如汉语介词 P 约有60个，连词 C约有110个，即状态 P 和 C分别对应的输出符号数为60、110。

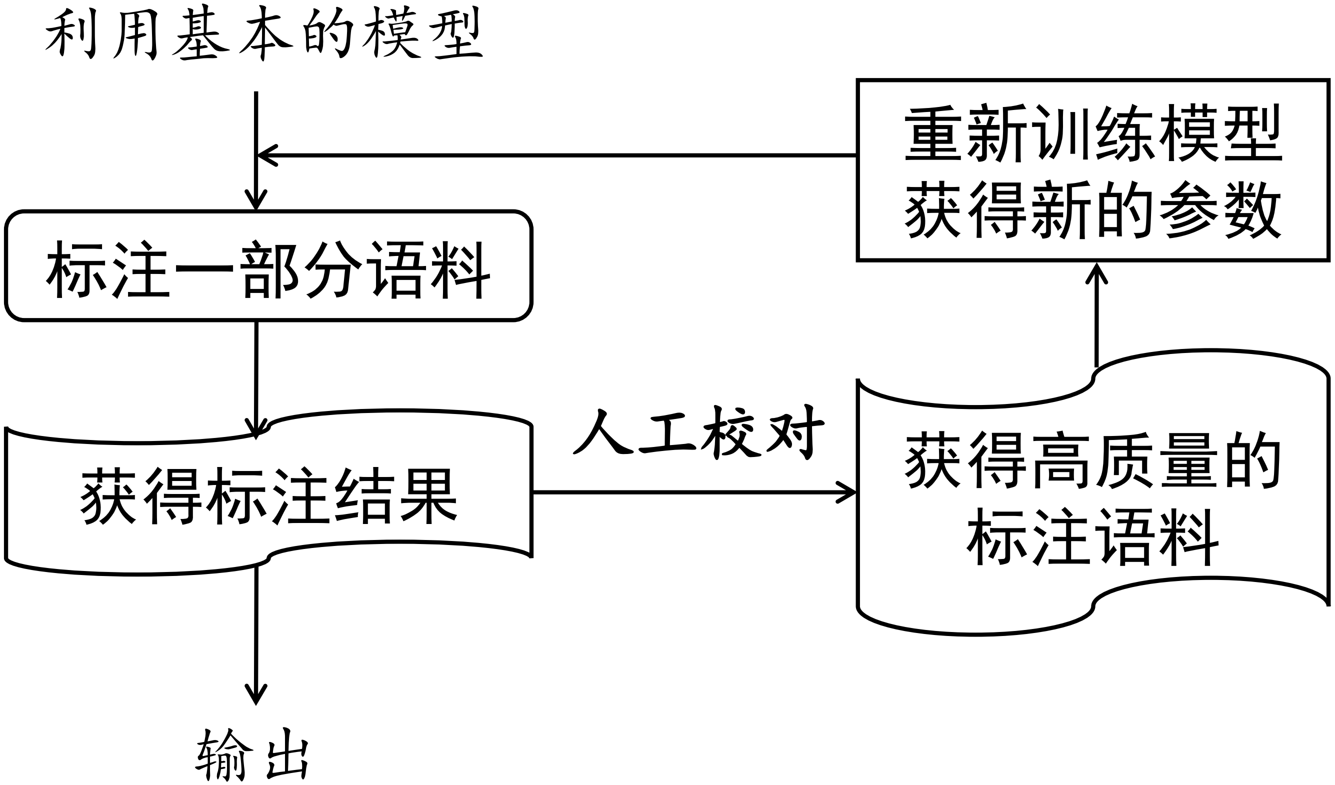
3. 隐马模型应用

- **参数估计**
  - 如果无任何标注语料，需要一部有词性标注的词典；
  - 如果有大量标注的语料，可从标注语料中获取词类个数；
  - 获取对应每种词类的词汇数(输出符号数)；
  - 基于标注语料统计，或者利用 EM 迭代算法获取初始状态概率、状态转移概率和输出符号概率。

通常情况下，在标注语料（训练集）上进行参数训练，在开发集上进行参数优化。或者利用初步训练的模型进行语料标注、校对，扩大训练集。

3. 隐马模型应用

一般地，需要通过错误驱动的机器学习方法修正模型的参数：



3. 隐马模型应用

- **北京大学分词和词性标注语料**

咱们/rr 中国/ns 这么/rz 大{da4}/a 的{de5}/ud 一个/mq 多/a 民族/n 的{de5}/ud 国家/n 如果/c 不/df 团结/a ，/wd 就/d 不/df 可能/vu 发展/v 经济/n ，/wd 人民/n 生活/n 水平/n 也/d 就/d 不/df 可能/vu 得到/v 改善/vn 和{he2}/c 提高/vn 。/wj

$$\bar{\pi}_{\mathrm{pos}_i} = \frac{\mathrm{POS}_i\text{出现在句首的次数}}{\text{所有句首的个数}}$$

$$\bar{a}_{ij} = \frac{\text{从词类}\mathrm{POS}_i\text{转移到}\mathrm{POS}_j\text{的次数}}{\text{所有从状态}\mathrm{POS}_i\text{转移到另一}\mathrm{POS}(\text{包括}\mathrm{POS}_j)\text{的总数}}$$

$$\bar{b}_j(k) = \frac{\text{从状态}\mathrm{POS}_j\text{输出词汇}w_k\text{的次数}}{\text{状态}\mathrm{POS}_j\text{出现的总次数}}$$

3. 隐马模型应用

- **分词性能：**

  (1)封闭测试：《人民日报》1998年1月份的部分切分和标注语料，约占训练语料的1/10，共78396个词，含中国人名1273个。(人名识别前)准确率：90.34%。

  (2)开放测试：《人民日报》1998年2月份的部分切分和标注语料，也占训练语料的1/10，共82347个词，含中国人名2316个。(人名识别前)准确率：86.32%。

汉语自动分词和中文人名识别技术研究, XX大学硕士学位论文, 2006

3. 隐马模型应用

- **词性标注：**

  (1)训练语料：北京大学标注的《人民日报》2000年1、2、4月份的语料；

  (2)封闭测试：2000年2月20~29日的标注语料，词性标注的精确率为：95.16%；

  (3)开放测试：2000年3月1~7日的语料，词性标注的精确率为：88.45%。

3. 隐马模型应用

- **训练语料规模对模型参数的影响：**

选用北大标注的2000年《人民日报》语料作为训练数据。5个训练语料集大小不同: C1为2月份的；C2为1月及2月份的；C3为1、2和4月份的；C4为1、2、4和9月份的；C5为1、2、4、9和10月份五个月的。采用相同的测试集(2000年3月份前7天的语料)，观察词性标注的精确率变化：

语料	C1	C2	C3	C4	C5
精确率(%)	86.16	90.85 ?	88.45	88.82	89.04

应用于词性标注的隐马尔可夫模型参数估计，YY大学硕士学位论文，2006

本章小结

◆马尔科夫模型

◆HMM 的构成

五元组：①状态数 ②输出符号数 ③初始状态的概率分布 ④状态转移的概率 ⑤输出概率

◆HMM 的三个基本问题

(1)快速计算给定模型的观察序列概率: 前/后向算法

(2)求最优状态序列：Viterbi 算法

(3)参数估计：Baum-Welch 算法

◆HMM在NLP中的应用

以汉语分词为例。

本章内容

1. 马尔科夫模型
2. 隐马尔可夫模型
3. 隐马模型应用
4. 习题

4. 习题

1. 请下载 HTK 开源代码，调试运行，体会该工具的使用方法。
2. 利用北京大学计算语言学研究所开放的《人民日报》标注语料，借助HTK工具实现汉语分词与词性标注方法。

谢谢！

Thanks!