隐马尔科夫模型（HMM）

哈尔滨工业大学计算学部 刘远超

隐马尔科夫模型(HMM)

本章内容简介

- HMM的基本概念
- HMM的概率计算算法

隐马尔科夫模型

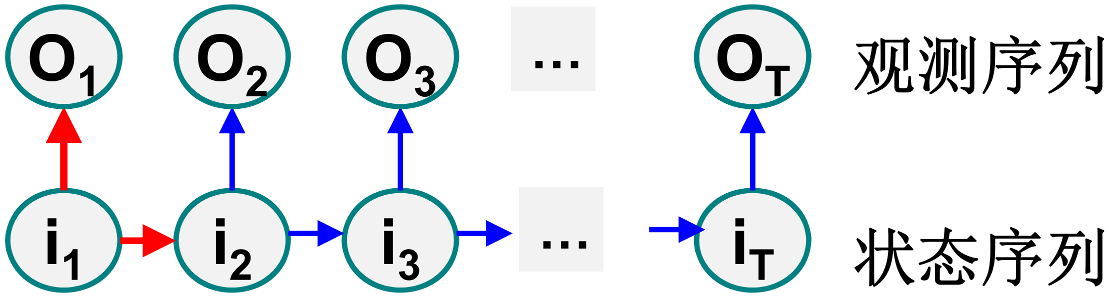
本章内容简介

- **HMM的基本概念**
- HMM的概率计算算法

隐马尔科夫模型

隐马尔科夫模型的基本概念

■隐马尔可夫模型（Hidden Markov Model，HMM）是关于时序的概率模型:



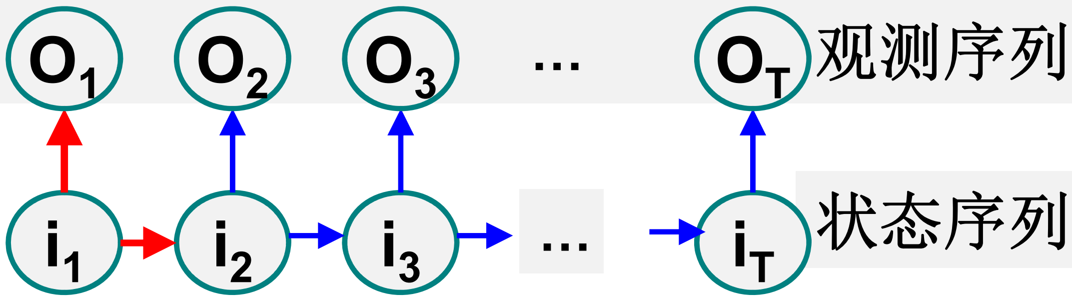
■ 描述：

1）由一个隐藏的马尔可夫链随机生成不可观测的状态随机序列(state sequence)；

2）再由各个状态生成一个观测而产生观测随机序列(observation sequence)的过程。

■ 序列的每一个位置也可以被视为是一个时刻。

隐马尔科夫模型HMM的基本概念



初始状态概率分布π

状态转移矩阵**A**

观测概率矩阵**B**

■HMM的形式定义如下：

- 设Q为所有可能状态的集合，V为所有可能观测的集合

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

  其中N为可能的状态数，M为可能的观测数。

- 设I为长度为T的状态序列，设O为对应的观测序列：

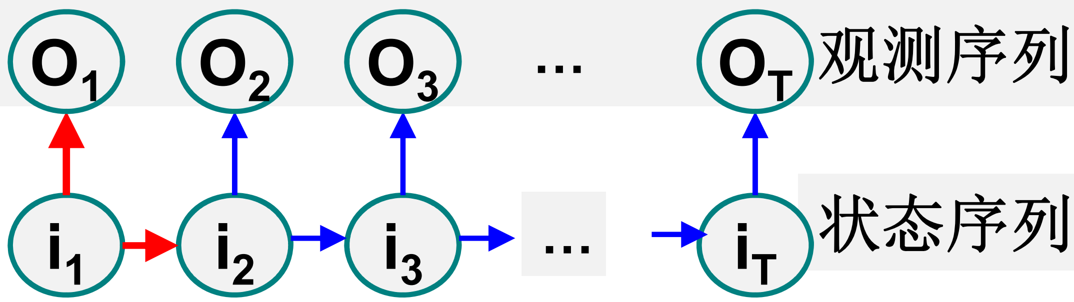
  $$\boldsymbol{I}=(\boldsymbol{i}_1,\boldsymbol{i}_2,\dots,\boldsymbol{i}_T), \boldsymbol{O}=(\boldsymbol{o}_1,\boldsymbol{o}_2,\dots,\boldsymbol{o}_T)。$$

- 设**A**为状态转移概率矩阵

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

其中，$a_{ij}=P(i_{t+1}=q_j|i_t=q_i)$,$i=1,2,\dots,N;\ j=1,2,\dots,N$是时刻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.

隐马尔科夫模型HMM的基本概念



初始状态概率分布π

状态转移矩阵A

观测概率矩阵B

● 设**B**为观测概率矩阵

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

其中$b_j(k) = P\left(o_t = v_k \middle| i_t = q_j\right), k = 1,2,\dots,M;\ j = 1,2,\dots,N$

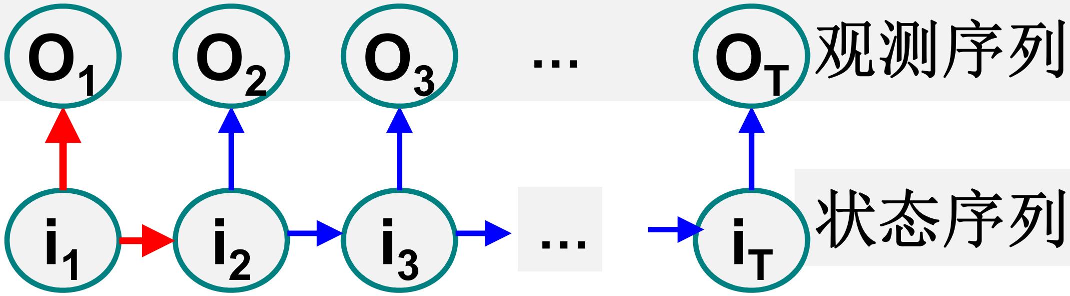
● 设$\boldsymbol{\pi}$是初始状态概率向量

$$\pi = (\pi_i)$$

其中，$\pi_i = \mathrm{P}(i_1 = q_i), \mathrm{i} = 1,2,\dots,\mathrm{N}$是时刻t=1处于状态$q_i$的概率

隐马尔科夫模型

隐马尔科夫模型HMM的基本概念

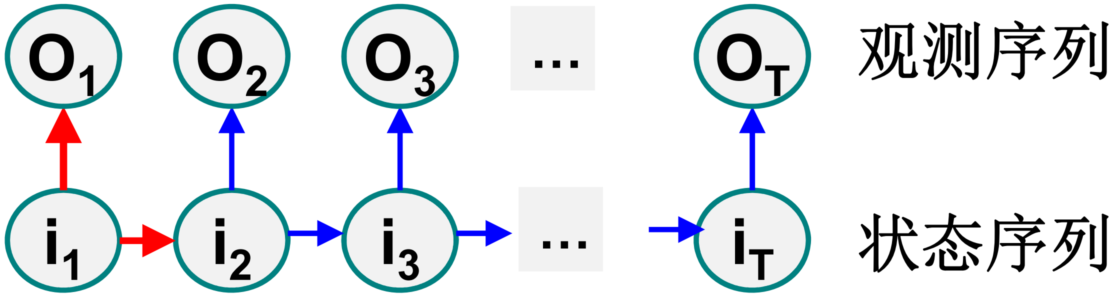


初始状态概率分布π
状态转移矩阵A
观测概率矩阵B

- 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（隐藏的HMM链），B决定观测序列。
- 因此，HMM模型λ可以用三元符号表示，即λ=（A, B, π）,成为HMM的三要素。

隐马尔科夫模型

隐马尔科夫模型HMM的基本概念



- HMM的两个基本假设：
  - **齐次马尔科夫性假设**：假设隐马尔可夫链t时刻的状态只和$t-1$时刻的状态有关：$P(i_t|i_{t-1},o_{t-1},\dots,i_1,o_1)=P(i_t|i_{t-1}),\ t=1,2,\dots,T$
  - **观测独立性假设**，假设观测只和当前时刻状态有关:

$$P(o_t|i_T,o_T,i_{T-1},o_{T-1},\dots,i_{t+1},o_{t+1},i_t,i_{t-1},o_{t-1},\dots,i_1,o_1)=P(o_t|i_t)$$

假设有4个盒子，每个盒子里都装有红白两种颜色的球，盒子里的红白球数量如下表

可以得到观测概率矩阵B

	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：

初始状态概率分布π

- 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；

状态转移矩阵A

- 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：

$$O = (红, 红, 白, 白, 红)$$

在这个过程中，观察者只能观测到球的颜色的序列，观测不到球是从哪个盒子取出的，即观测不到盒子的序列。

盒子和球模型

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

- 状态集合：Q={盒子1，盒子2，盒子3，盒子4}，N=4（可能的状态数）
- 观测集合：V={红球，白球} M=2（可能的观测数）
- 初始化状态概率分布：

$$\pi = (0.25, 0.25, 0.25, 0.25)^{\mathrm{T}}$$

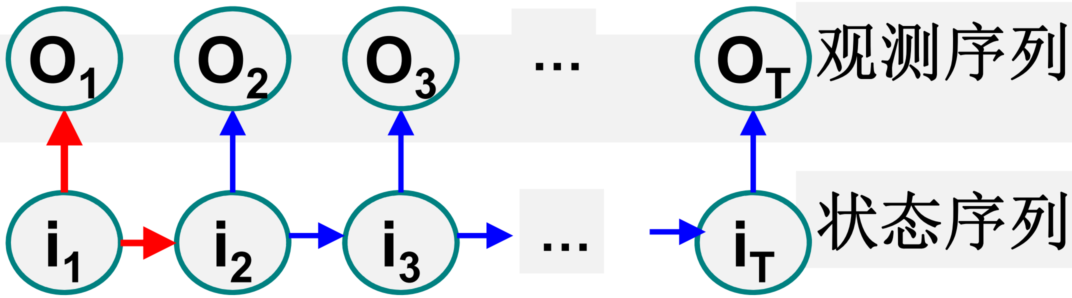
- 状态转移矩阵A：

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0.4 & 0 & 0.6 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \end{bmatrix}$$

观测概率矩阵**B**：

$$B = \begin{bmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \\ 0.6 & 0.4 \\ 0.8 & 0.2 \end{bmatrix}$$

HMM观测序列的生成过程



初始状态概率分布π

状态转移矩阵**A**

观测概率矩阵**B**

根据隐马尔可夫模型定义，可以将一个长度为T的观测序列$O=(o_1,o_2,\cdots,o_T)$的生成过程描述如下：

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$t=1$

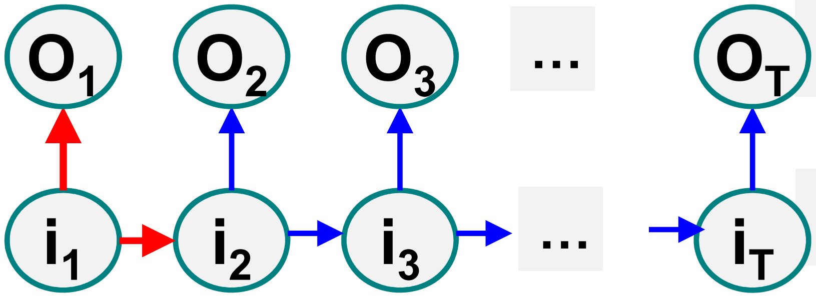
（3）按照状态i_t的观测概率分布$b_{i_t}(k)$生成o_t

（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 ■

隐马尔科夫模型

隐马尔科夫模型的三个基本问题



观测序列

状态序列

初始状态概率分布π

状态转移矩阵A

观测概率矩阵B

（1）概率计算问题。给定模型$\lambda=$（A,B,π）和观测序列$O=(O_1,O_2,\dots,O_T)$，计算在模型λ下观测序列O出现的概率$P(O|\lambda)$

（2）学习问题。已知观测序列$O=(O_1,O_2,\dots,O_T)$，估计模型$\lambda=$（A,B,π）的参数，使得在该模型下观测序列O出现的概率$P(O|\lambda)$最大，即用极大似然估计的方法估计参数

（3）预测问题。也称为解码（decoding）问题。已知模型$\lambda=$（A,B,π）和观测序列$O=(o_1,o_2,\dots,o_T)$，求对给定观测序列条件概率$P(I|O)$最大的状态序列$I=(i_1,i_2,\dots,i_T)$，即给定观测序列，求最有可能的对应的状态序列。

隐马尔科夫模型

本章内容简介

- HMM的基本概念
- HMM的概率计算算法

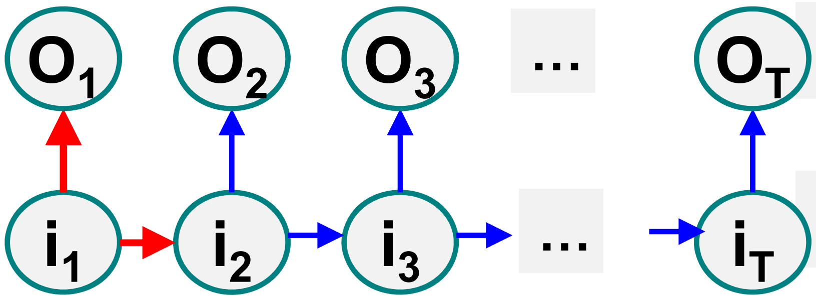
隐马尔科夫模型

本章内容简介

- HMM的基本概念
- **HMM的概率计算算法**

计算观测序列概率$P(O|\lambda)$的算法

（1）先介绍概念上可行但实际不可行的直接计算法



观测序列

状态序列

初始状态概率分布$\boldsymbol{\pi}$

状态转移矩阵A

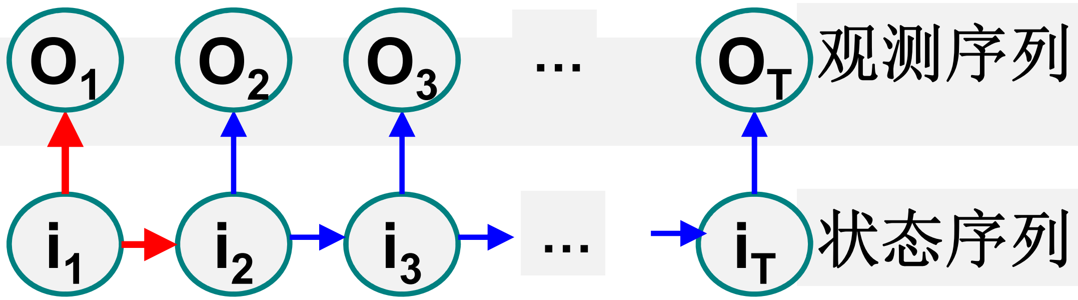
观测概率矩阵B

概率计算问题的问题描述：给定模型$\boldsymbol{\lambda}=$（$\boldsymbol{A},\boldsymbol{B},\boldsymbol{\pi}$）和观测序列$\boldsymbol{O}=(\boldsymbol{O_1},\boldsymbol{O_2},\ldots,\boldsymbol{O_T})$,计算在模型$\boldsymbol{\lambda}$下观测序列O出现的概率$\mathbf{P}(\boldsymbol{O}|\boldsymbol{\lambda})$

最直接的方法：

1. 列举所有可能的长度为T的状态序列$\boldsymbol{I}=(\boldsymbol{i_1},\boldsymbol{i_2},\ldots,\boldsymbol{i_T})$;
2. 求各个状态序列 *I* 与观测序列 $\boldsymbol{O}=(\boldsymbol{O_1},\boldsymbol{O_2},\ldots,\boldsymbol{O_T})$的联合概率$P(O,I|\lambda)$;
3. 然后对所有可能的状态序列求和，得到$P(O|\lambda)$.

（1）概念上可行但实际不可行的直接计算观测序列概率$P(O|\lambda)$法



初始状态概率分布$\boldsymbol{\pi}$

状态转移矩阵**A**

观测概率矩阵**B**

1. 状态序列$\boldsymbol{I} = (\boldsymbol{i_1}, \boldsymbol{i_2}, \dots, \boldsymbol{i_T})$的概率是 $\boldsymbol{P(I|\lambda) = \pi_{i_1} a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_{T-1} i_T}}$ ；

2. 对每个固定的状态序列$\boldsymbol{I} = (\boldsymbol{i_1}, \boldsymbol{i_2}, \dots, \boldsymbol{i_T})$，观测序列 $\boldsymbol{O} = (\boldsymbol{O_1}, \boldsymbol{O_2}, \dots, \boldsymbol{O_T})$的概率是$P(O|I，\lambda)$：

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

3. O 和 I同时出现的联合概率为

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

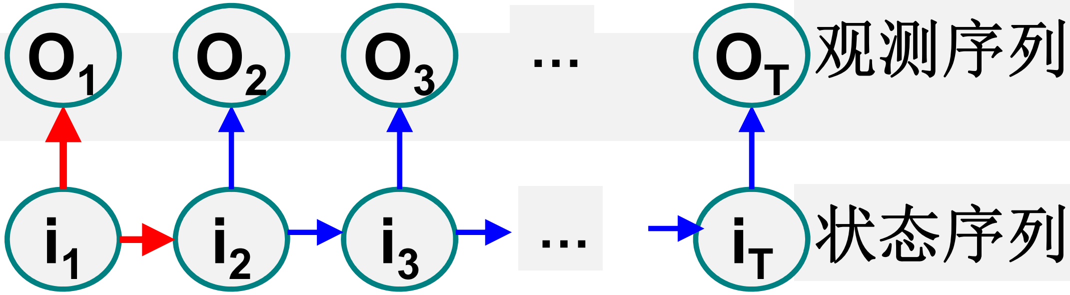
4. 然后，对所有可能的状态I 求和， 得到观测序列O的概率$P(O|\lambda)$，即

$$P(O|\lambda) = \sum_I P(O|I,\lambda)\boldsymbol{P(I|\lambda)}$$

$$= \sum_{i_1, i_2, \dots, i_T} \boldsymbol{\pi_{i_1}} b_{i_1}(o_1) \boldsymbol{a_{i_1 i_2}} b_{i_2}(o_2) \dots \boldsymbol{a_{i_{T-1} i_T}} b_{i_T}(o_T)。$$

但该公式计算量很大，为$O(TN^T)$，不可行

计算观测序列概率$P(O|\lambda)$的有效算法:前向算法



初始状态概率分布$\boldsymbol{\pi}$
状态转移矩阵**A**
观测概率矩阵**B**

■ 首先定义**前向概率**：给定隐马尔科夫模型λ，定义到时刻t的部分观测序列为：$\boldsymbol{o_1, o_2, \ldots, o_t}$，且状态为$q_i$的概率为前向概率，记作：$\alpha_t(i) = P(\boldsymbol{o_1, o_2, \ldots, o_t}, i_t = q_i|\lambda)$

■ 可以递推地求得前向概率$\alpha_t(i)$及**观测序列概率**$\mathbf{P}(\boldsymbol{O}|\boldsymbol{\lambda})$

■ **观测序列概率**$\mathbf{P}(\boldsymbol{O}|\boldsymbol{\lambda})$的前向算法

● 输入：隐马尔科夫模型λ，观测序列0;

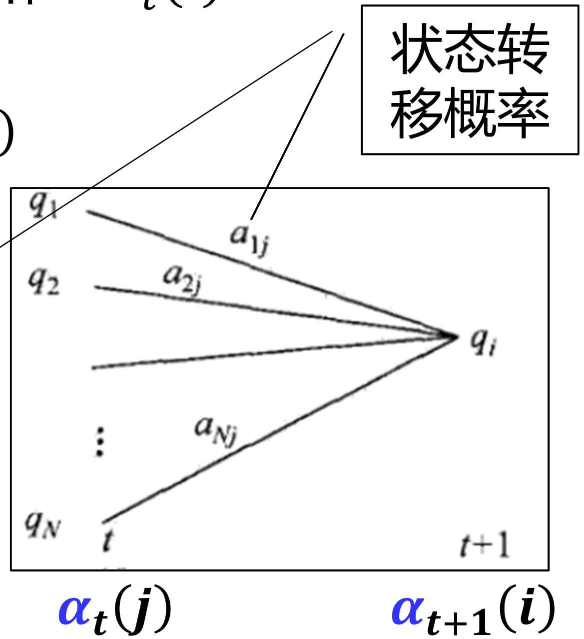
● 输出：**序列概率**$\mathbf{P}(\boldsymbol{O}|\boldsymbol{\lambda})$

（1）初始值：$\alpha_1(i) = \pi_{\mathrm{i}} b_i(o_1),\ i = 1,2, \ldots, N;$

(2) 递推：对$\mathrm{t} = 1,2, \ldots, \mathrm{T}-1,$

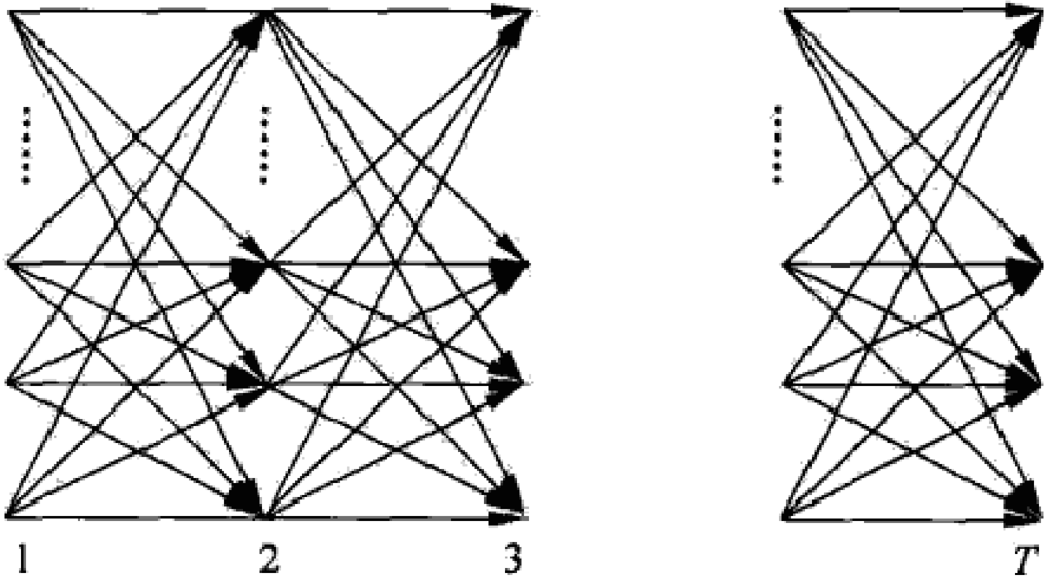
$$\alpha_{t+1}(i) = \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji}\right] b_i(o_{t+1}),\ i = 1,2, \ldots, N;$$

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



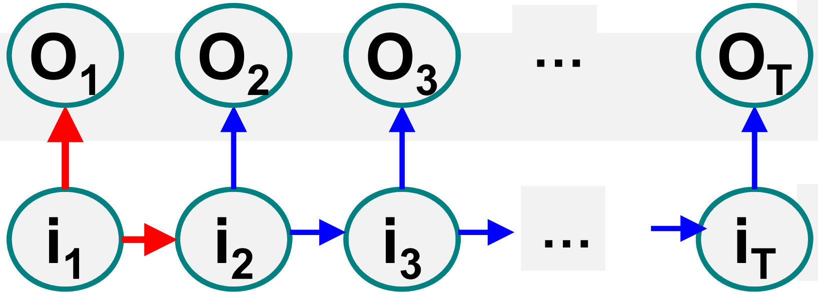
计算观测序列概率$P(O|\lambda)$的有效算法:前向算法

前向算法的观测序列结构



- 每个$\alpha_{t+1}(i)$的计算利用前一时刻的N个$\alpha_t(j)$，因此前向算法减少计算量的原因在于每一次计算，直接引用前一个时刻的计算结果，避免重复计算。计算复杂度是O(N^2T)；

计算观测序列概率$P(O|\lambda)$的有效算法:前向算法



观测序列

状态序列

初始状态概率分布$\boldsymbol{\pi}$

状态转移矩阵**A**

观测概率矩阵**B**

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

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

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

解：按照算法,（1）计算初值 $\boxed{\alpha_1(i) = \pi_i b_i(o_1), i = 1,2,\ldots,N;}$

$$\alpha_1(1) = \pi_1 b_1(o_1) = 0.10$$

$$\alpha_1(2) = \pi_2 b_2(o_1) = 0.16$$

$$\alpha_1(3) = \pi_3 b_3(o_1) = 0.28$$

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

（2）递推：对$\mathrm{t}=1,2,\ldots,\mathrm{T}-1$,

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

状态转移概率

$$\alpha_2(1)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i1}\right]b_1(o_2)=0.154\times0.5=0.077$$

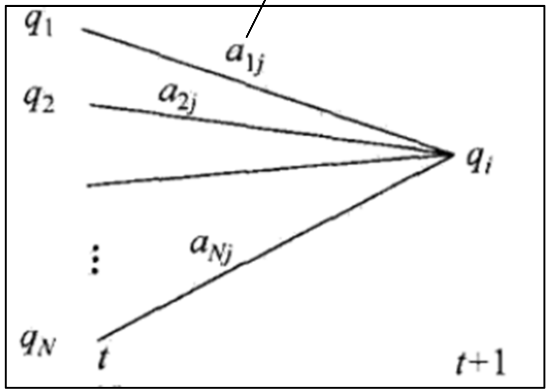
$$\alpha_2(2)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i2}\right]b_2(o_2)=0.184\times0.6=0.1104$$

$$\alpha_2(3)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i3}\right]b_3(o_2)=0.202\times0.3=0.0606$$

$$\alpha_3(1)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i1}\right]b_1(o_3)=0.04187$$

$$\alpha_3(2)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i2}\right]b_2(o_3)=0.03551$$

$$\alpha_3(3)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i3}\right]b_3(o_3)=0.05284$$



$\alpha_t(j)$ $\alpha_{t+1}(i)$

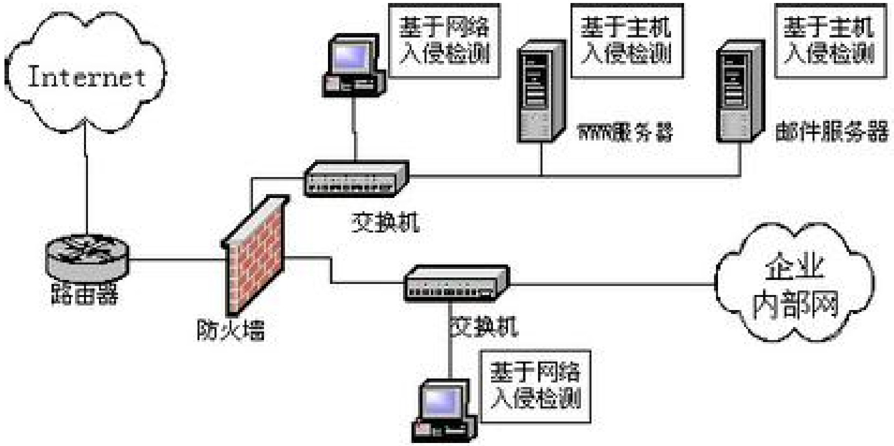
（3）终止

$$P(O\mid\lambda)=\sum_{i=1}^{3}\alpha_3(i)=0.13022$$

隐马尔科夫模型

HMM的应用场景

- 人脸识别
- 语音识别
- 入侵检测



本章小节

- HMM的基本概念
- HMM的概率计算算法