NLP基础2023春

HMM模型数学推导-补充

主讲人：明训

- HMM 要解决的问题
- 已知模型 HMM 用 λ 表示， λ = (π, A, B)
- **在给定模型 λ 的情况下，出现了观测序列 O = {o1, o2, o3,...,oT}，计算该序列出现的概率 P(O | λ)**
- 在给定观测序列 O = {o1, o2, o3,...,ot} 的情况下， 求解参数 λ = (π, A, B), 使得观测序列 O 出现的概率最大 $\lambda_{MLE} = argmax_{\lambda} P(O|\lambda)$
- 在给定模型 λ 的情况下， 出现了观测序列 O = {o1, o2, o3,...,oT}， 求隐藏序列 I， 使得P(I | O, λ) 的概率最大 $\hat{I} = argmax_{I} P(I|O)$

Hidden Markov 概念

DT ACADEMICS

- 初始概率分布 π = $\{\pi_1, \pi_2, .. \pi_P\}$
- 隐藏状态序列 I = $\{i_1, i_2, .. i_T\}$
- 隐藏状态合集 Q = $\{q_1, q_2, .. q_N\}$
- 观察状态序列 O = $\{o_1, o_2, .. o_T\}$
- 观测状态合集 V = $\{v_1, v_2, .. v_M\}$
- 隐藏状态转移矩阵 A $a_{i,j} = P(i_{t+1} = q_j | i_t = q_i)$
- 观测状态转移矩阵 B $b_j(k) = P(o_t = v_k | i_t = q_j)$

$$\begin{array}{ccc} & & \textit{Today} \\ & & \begin{array}{ccc} \textit{sun} & \textit{cloud} & \textit{rain} \end{array} \\ \textit{Yesterday} & \begin{array}{c} \textit{sun} \\ \textit{cloud} \\ \textit{rain} \end{array} & \begin{bmatrix} 0.50 & 0.375 & 0.125 \\ 0.25 & 0.125 & 0.625 \\ 0.25 & 0.375 & 0.375 \end{bmatrix} \end{array}$$

	海藻状态				
		干燥	半干	湿润	湿透
天气状态	晴天	0.6	0.2	0.15	0.05
	多云	0.25	0.25	0.25	0.25
	雨天	0.05	0.1	0.35	0.5

Hidden Markov 概念

- **在给定模型 λ 的情况下，出现了观测序列 O = {o1, o2, o3,...,oT}，计算该序列出现的概率 P(O | λ)**
- **Brute Force**
- 在指定模型下，寻找指定观测序列出现的概率 = 指定模型下，寻找所有可能的隐藏状态序列产生指定的观测序列出现的概率

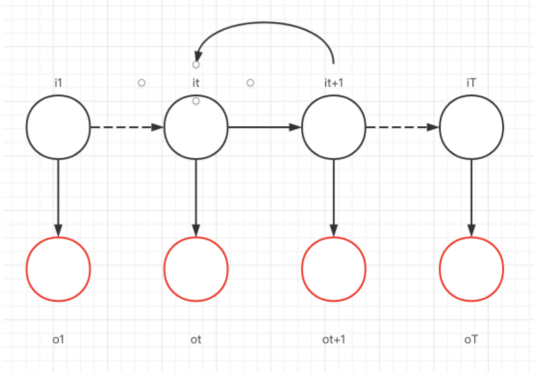
$$P(O|\lambda) = \sum_{I} P(O, I|\lambda)$$

$$P(O, I|\lambda) = P(I|\lambda)P(O|I, \lambda)$$

Hidden Markov 概念

Hidden Markov 概念

- P(I | λ) = 在给定模型 λ 的前提下，隐藏状态序列 I 出现的概率



$$I = \{i_1, i_2, \ldots, i_T\}$$

$$P(I|\lambda) = P(i_1, i_2 \ldots, i_T|\lambda)$$

$$P(i_1, i_2 \ldots, i_T|\lambda) = P(i_1|\lambda)P(i_2, \ldots i_T|i_1, \lambda)$$

$$P(i_1, i_2 \ldots, i_T|\lambda) = P(i_1|\lambda)P(i_2, i_3 \ldots i_T|i_1, \lambda) = P(i_1|\lambda)P(i_2|i_1, \lambda)P(i_3, i_4, \ldots i_T|i_1, i_2, \lambda)$$

$$P(i_1|\lambda)P(i_2|i_1, \lambda)P(i_3, i_4, \ldots i_T|i_1, i_2, \lambda) = P(i_1|\lambda)P(i_2|i_1, \lambda) \ldots P(i_T|i_1, i_2, \ldots i_{T-1}, \lambda)$$

基于齐次马尔科夫假设

$$P(i_1|\lambda)P(i_2|i_1, \lambda) \ldots P(i_T|i_1, i_2, \ldots i_{T-1}, \lambda) = P(i_1)P(i_2|i_1)P(i_3|i_2) \ldots P(i_T|i_{T-1})$$

$$P(i_1)P(i_2|i_1)P(i_3|i_2) \ldots P(i_T|i_{T-1}) = a_{i_{T-1,T}} a_{i_{T-2,T-1}} a_{i_{1,2}} \pi_i a_{i_1}$$

$$a_{i_{t-1,t}} a_{i_{t-2,t-1}} a_{i_{1,2}} \pi_i a_{i_1} = \pi_i a_{i_1} \prod_{t=2}^{T} a_{i_{t-1,t}}$$

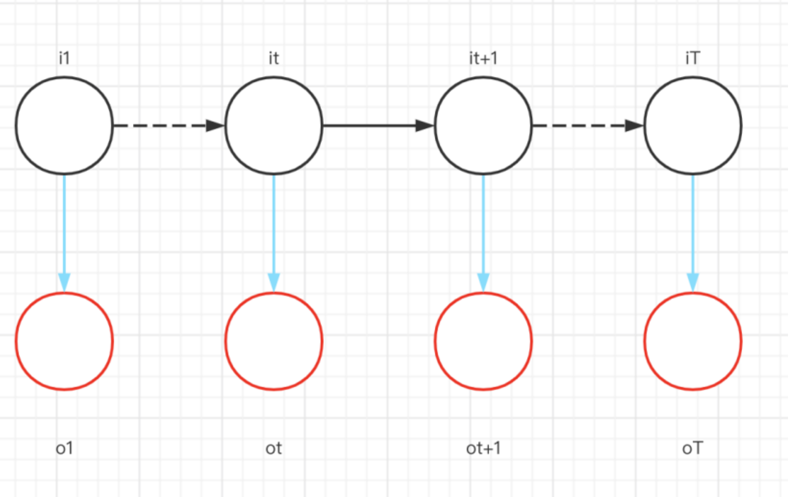
$$\begin{array}{cc} & \textit{Today} \\ & \begin{array}{ccc} \textit{sun} & \textit{cloud} & \textit{rain} \end{array} \\ \textit{Yesterday} \begin{array}{r} \textit{sun} \\ \textit{cloud} \\ \textit{rain} \end{array} & \begin{bmatrix} 0.50 & 0.375 & 0.125 \\ 0.25 & 0.125 & 0.625 \\ 0.25 & 0.375 & 0.375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

- P(O| I, λ) = 在给定模型 λ和隐藏状态序列 I 的前提下，出现观测序列 O= $\{o_1, o_2, ..o_T\}$的概率
- P(O |I, λ) = P(O | I) = $P(o_1, o_2.., o_T|i_1, i_2.., i_T)$
- 基于观测独立性假设

$$P(o_1, o_2.., o_T|i_1, i_2.., i_T) = P(o_1|i_1, ...i_t)P(o_2, ..o_T|i_1, ..i_T) = b_{i_1}(o_1)b_{i_2}(o_2)...b_{i_T}(o_T)$$

$$b_{i_1}(o1)b_{i_2}(o2)...b_{i_t}(o_t) = \prod_{t=1}^{T} b_{i_t}(o_t)$$

		海藻状态			
		干燥	半干	湿润	湿透
天气状态	晴天	0.6	0.2	0.15	0.05
	多云	0.25	0.25	0.25	0.25
	雨天	0.05	0.1	0.35	0.5



Hidden Markov 概念

$$P(O,I|\lambda) = P(I|\lambda)P(O|I,\lambda)$$

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

$$P(I|\lambda)P(O|I,\lambda) = \sum_{I} \pi(a_{i_1}) \prod_{t=2}^{T} a_{i_{t-1},i_t} \prod_{t=1}^{T} b_{i_t}(o_t)$$

$$\sum_{I} = \sum_{i_1,i_2,\ldots,i_T}$$

每一个隐藏状态位置都有N种可能性，所以这里的时间复杂度是 $O(N^T)$

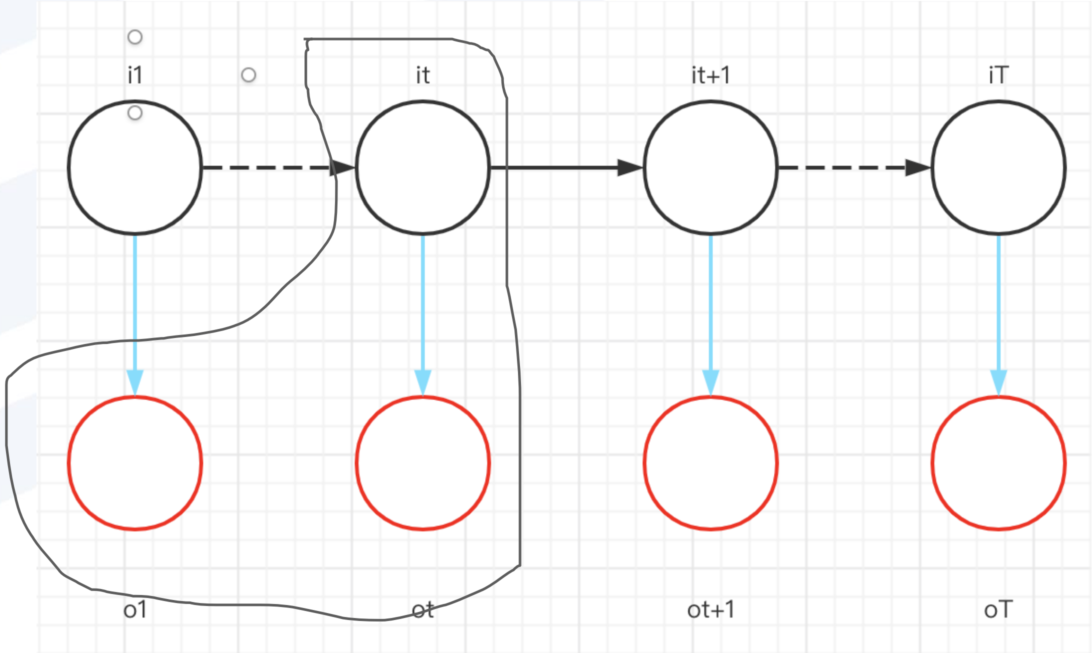
Brute Force 是一种解决思路，但肯定不是好解决思路

- 前向算法 Forward
- 给定t时刻的隐藏状态为i，观测序列 $\{o_1, o_2 \ldots, o_t\}$ 的概率

$$\alpha_t(i) = P(o_1, o_2 \ldots, o_t, i_t = q_i|\lambda)$$

$$\alpha_T(i) = P(o1, \ldots oT, i_T = q_i|\lambda) = P(O, i_T = q_i|\lambda)$$

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



$$\alpha_{t+1}(j) = P(o_1, o_2 \ldots, o_{t+1}, i_{t+1} = q_j|\lambda)$$

$$P(o_1, o_2 \ldots, o_{t+1}, i_{t+1} = q_j|\lambda) = \sum_{i=1}^{N} P(o_1, \ldots o_{t+1}, i_{t+1} = q_j, i_t = q_i|\lambda)$$

$$\sum_{i=1}^{N} P(o1, \ldots o_{t+1}, i_{t+1} = q_j, i_t = q_i|\lambda) = \sum_{i=1}^{N} P(o_{t+1}|o1, \ldots o_t, i_{t+1} = q_j, i_t = q_i, \lambda) P(o_1, \ldots o_t, i_t = q_i, i_{t+1} = q_j|\lambda)$$

$$\sum_{i=1}^{N} P(o_{t+1}|o1, \ldots o_t, i_{t+1} = q_j, i_t = q_i, \lambda) P(o_1, \ldots o_t, i_t = q_i, i_{t+1} = q_j|\lambda) = \sum_{i=1}^{N} P(o_{t+1}|i_{t+1} = q_j, \lambda) P(o_1, \ldots o_t, i_t = q_i, i_{t+1} = q_j$$

$$= \sum_{i=1}^{N} P(o_{t+1}|i_{t+1} = q_j|\lambda) P(i_{t+1} = q_j|o_1, ..o_t, i_t = q_i, \lambda) P(o_1, ..o_t, i_t = q_i|\lambda)$$

$$= \sum_{i=1}^{N} P(o_{t+1}|i_{t+1}) P(i_{t+1} = q_j|i_t = q_i) \alpha_t(i) = \sum_{i=1}^{N} b_j(O_{t+1}) a_{i,j} \alpha_t(i)$$

Hidden Markov 概念

- 假设有三个盒子，每个盒子有红白两种颜色的球
- 盒子1：5红5白
- 盒子2：4红6白
- 盒子3：7红3白

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

- 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
- 观测序列结果：O = {红，白，红}

Hidden Markov 概念

- 时刻1
- （红色球，盒子1） = $\alpha_1(1) = \pi_1 b_1(o_1) = 0.2 * 0.5 = 0.1$
- （红色球，盒子2） = $\alpha_1(2) = \pi_2 b_2(o_1) = 0.4 * 0.4 = 0.16$
- （红色球，盒子3） = $\alpha_1(3) = \pi_3 b_3(o_1) = 0.4 * 0.7 = 0.28$

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

- 时刻2
- （白色球，盒子1）$=\alpha_2(1)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i_1}\right]b_1(o2)=[0.1*0.5+0.16*0.3+0.28*0.2]*0.5=0.077$
- （白色球，盒子2）$=\alpha_2(2)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i_2}\right]b_2(o2)=[0.1*0.2+0.16*0.5+0.28*0.3]*0.5=0.1104$
- （白色球，盒子3）$=\alpha_2(3)=\left[\sum_{i=1}^{3}\alpha_1(i)a_{i_3}\right]b_3(o2)=[0.1*0.3+0.16*0.2+0.28*0.5]*0.5=0.0606$

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

- 时刻3
- （红色球，盒子1） $=\alpha_3(1)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i_1}\right]b_1(o3)=[0.077*0.5+0.1104*0.3+0.0606*0.2]*0.5=0.0$
- （红色球，盒子2） $=\alpha_3(2)=\left[\sum_{i=1}^{3}\alpha_2(i)a_{i_2}\right]b_2(o3)=[0.077*0.2+0.1104*0.5+0.0606*0.3]*0.4=0.0$
- （红色球，盒子3） $=\alpha_3(3)=\left[\sum_{i=1}^{3}\alpha_3(i)a_{i_3}\right]b_3(o3)=[0.077*0.3+0.1104*0.2+0.0606*0.5]*0.7=0.0$

$$P(O|\lambda)=\sum_{i=1}^{3}\alpha_3(i)=0.130$$

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

- 后向算法 Backward
- 给定t时刻的隐藏状态为i， 和观测序列$\{o_{t+1}, o_{t+2}\cdots, o_T\}$的概率

$$\beta_t(i) = P(o_{t+1},..o_T|i_t = q_i, \lambda)$$

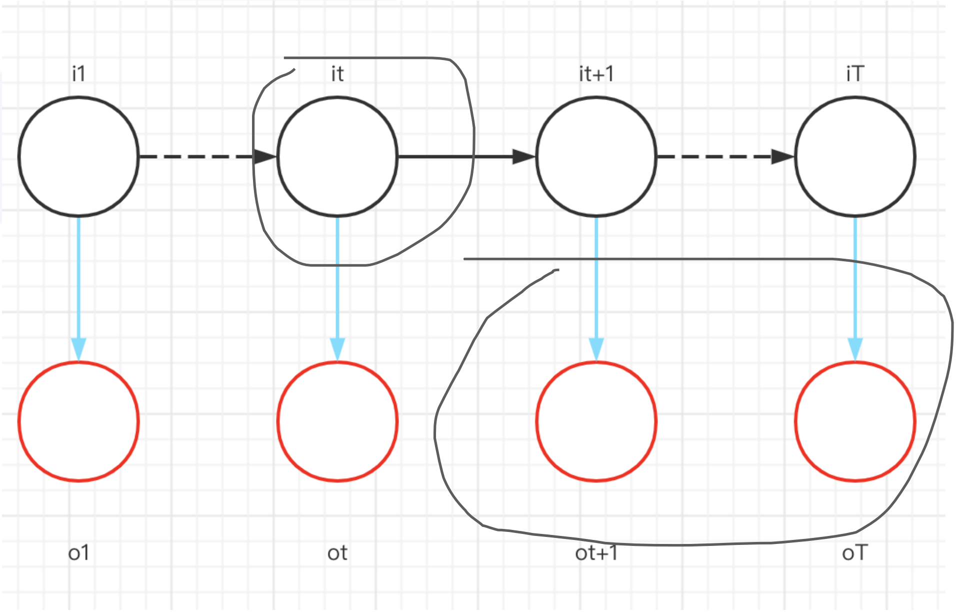
$$\beta_1(i) = P(o_2,..o_T|i_1 = q_i, \lambda)$$

$$P(O|\lambda) = P(o_1,..o_T|\lambda)$$

$$= \sum_{i=1}^{N} P(o_1,..,o_T, i_1 = q_i) = \sum_{i=1}^{N} P(o_1,..,o_T|i_1 = q_i)P(i_1 = q_i)$$

$$= \sum_{i=1}^{N} P(o_1|o_2,..,o_T, i_1 = q_i)P(o_2,..,o_T|i_1 = q_i)\pi_i$$

$$= \sum_{i=1}^{N} P(o_1|i_1 = q_i)\beta_1(i)\pi_i = \sum_{i=1}^{N} b_i(o_1)\beta_1(i)\pi_i$$



$$\beta_t(i) = P(O_{t+1}, \ldots, O_T | i_t = q_i)$$

$$= \sum_{j=1}^{N} P(O_{t+1}, .. O_T, i_{t+1} = q_j | i_t = q_i)$$

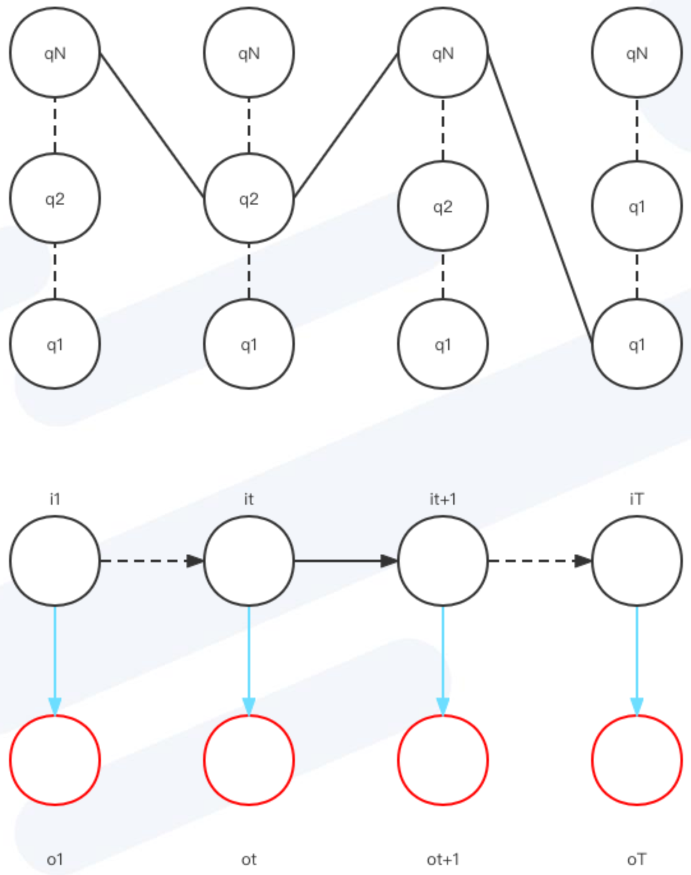
$$= \sum_{j=1}^{N} P(O_{t+1}, .. O_T | i_{t+1} = q_j, i_t = q_i) P(i_{t+1} = q_j | i_t = q_i)$$

$$= \sum_{j=1}^{N} P(O_{t+1}, .. O_T | i_{t+1} = q_j) a_{i,j}$$

$$= \sum_{j=1}^{N} P(o_{t+1} | o_{t+2}, .. O_T, i_{t+1} = q_j) P(o_{t+2}, .. o_T | i_{t+1} = q_j) a_{i,j}$$

$$= \sum_{j=1}^{N} P(o_{t+1} | i_{t+1} = q_j) \beta_{t+1}(j) a_{i,j}$$

- 在给定模型 λ 的情况下，出现了观测序列 O = {o1, o2, o3,…,oT}，求隐藏序列 I，使得P(I | O, λ) 的概率最大



Hidden Markov 概念

DT ACADEMICS

- 在给定模型 λ 的情况下，出现了观测序列 O = {o1, o2, o3,...,oT}，求隐藏序列 I，使得P(I | O, λ) 的概率最大
- t时刻的隐藏状态为qi的隐藏状态序列最大值为θ

$$\theta_t(i) = \max_{i_1,i_2,\dots i_{t-1}} P(o_1, o_2, \dots o_t, i_1, i_2, \dots, i_{t-1}, i_t = q_i, \lambda) = \max_{i_1,i_2,\dots i_{t-1}} P(o_1, o_2, \dots o_t, i_1, i_2, \dots, i_{t-1}, i_t = q_i)$$

$$\theta_{t+1}(j) = \max_{i_1,i_2,\dots i_t} P(o_1, o_2, \dots o_t, o_{t+1}, i_1, i_2, \dots, i_t, i_{t+1} = q_j)$$

$$= \max_{1 \le i \le N} \theta_t(i) a_{i,j} b_j(o_{t+1})$$

Hidden Markov 概念

DT ACADEMICS

- 维特比算法
- 在给定模型 λ 的情况下，出现了观测序列 O = {o1, o2, o3,...,oT}，求隐藏序列 I，使得P(I | O, λ) 的概率最大
- t时刻的隐藏状态为qi的隐藏状态序列最大值为θ

$$\delta_t(i) = \max_{i_1,i_2,\ldots i_{t-1}} P(o_1, o_2, \ldots o_t, i_1, i_2, \ldots, i_{t-1}, i_t = q_i, \lambda) = \max_{i_1,i_2,\ldots i_{t-1}} P(o_1, o_2, \ldots o_t, i_1, i_2, \ldots, i_{t-1}, i_t = q_i)$$

$$\delta_{t+1}(j) = \max_{i_1,i_2,\ldots i_t} P(o_1, o_2, \ldots o_t, o_{t+1}, i_1, i_2, \ldots, i_t, i_{t+1} = q_j)$$

$$= \max_{1 \le i \le N} \delta_t(i) a_{i,j} b_j(o_{t+1})$$

$$\varphi_{t+1}(j) = arg \max_{1 \le i \le N} \delta_t(i) a_{i,j}$$

- 隐状态转移矩阵

	Next		
Current	A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

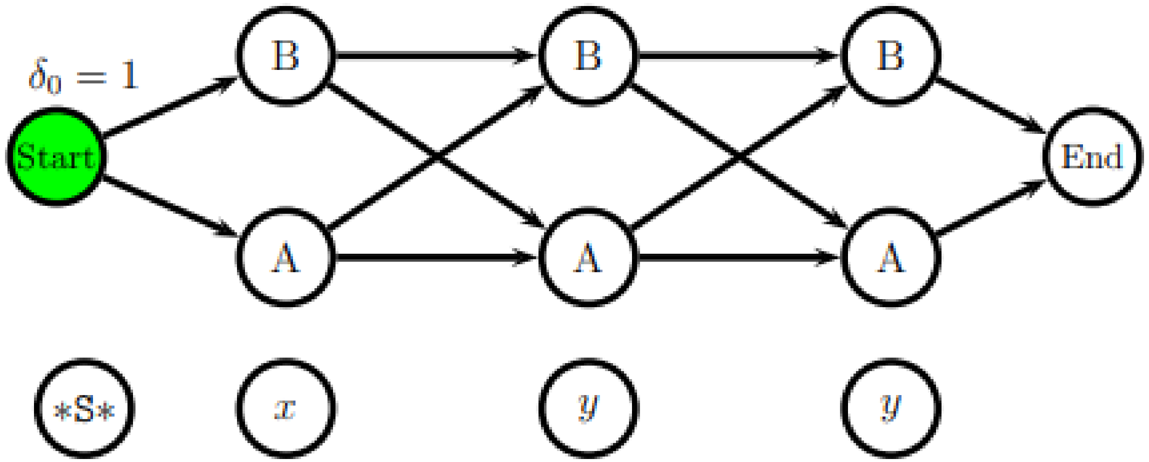
- 观测状态转移矩阵

	Word		
State	*S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

- 观测序列为 x y y
- 假设第0位置，隐藏序列取 <start sign> 的概率为1 $\delta_0(<start>) = 1$

Hidden Markov 概念

DT Academics



Current	Next A	Next B	Next End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

State	Word *S*	Word x	Word y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

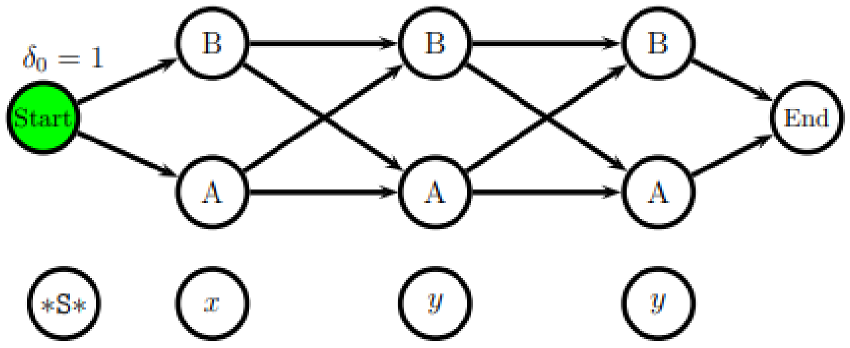
Hidden Markov 概念

$$\delta_1(A) = \max_{i_0} P(A|i_0)P(< start > |i_0)\delta_0(i_0)$$

$$\delta_1(A) = 1 \times 1 \times 0.7 = 0.7$$

$$\delta_1(B) = 1 \times 1 \times 0.3 = 0.3$$

$$\psi_1(A) = \psi_1(B) = O_0$$



Current	Next: A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

State	Word: *S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

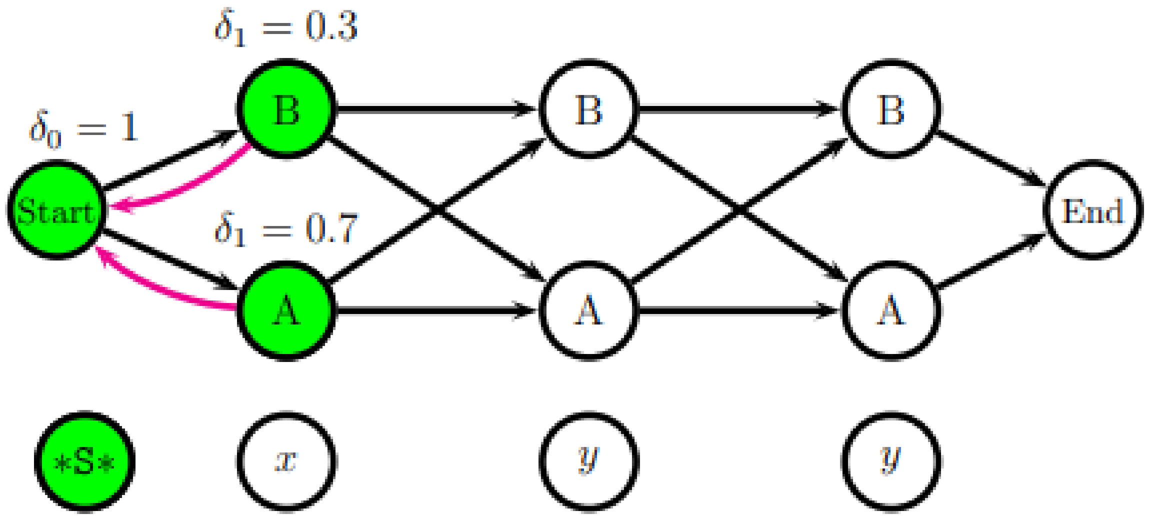
Hidden Markov 概念

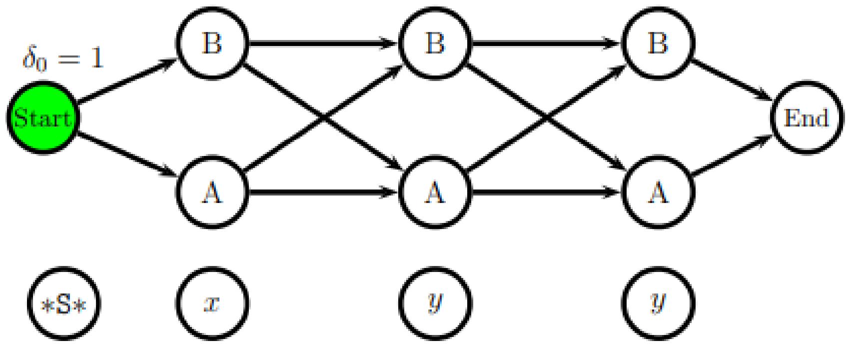
$$\delta_1(A) = \max_{i_0} P(A|i_0)P(*S*|i_0)\delta_0(i_0)$$

$$\delta_1(A) = 1 \times 1 \times 0.7 = 0.7$$

$$\delta_1(B) = 1 \times 1 \times 0.3 = 0.3$$

$$\psi_1(A) = \psi_1(B) = *S*_0$$





	Next		
Current	A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

	Word		
State	*S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

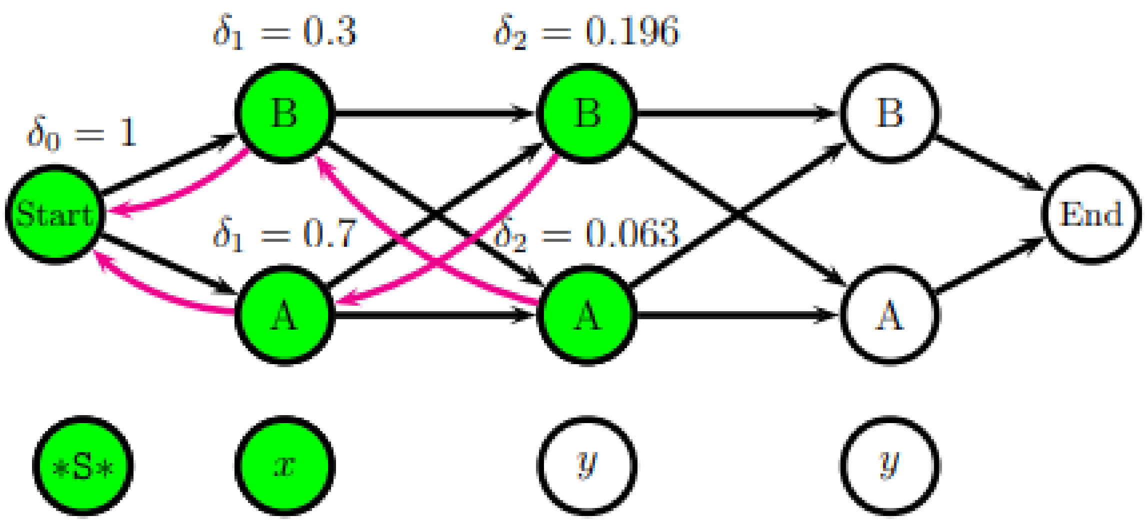
Hidden Markov 概念

DT ACADEMICS

$$\delta_2(A) = \max_{i_1} P(A|i_1)P(*S*|i_1)\delta_1(i_1) = max\{0.2 \times 0.4 \times 0.7, 0.7 \times 0.3 \times 0.3\} = 0.063$$

$$\psi_1(A) = B_1 \qquad \psi_1(B) = A_1$$

$$\delta_2(B) = \max_{i_1} P(B|i_1)P(*S*|i_1)\delta_1(i_1) = max\{\overbrace{0.7 \times 0.4 \times 0.7}, \overbrace{0.2 \times 0.3 \times 0.3}\} = 0.196$$



	Next		
Current	A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

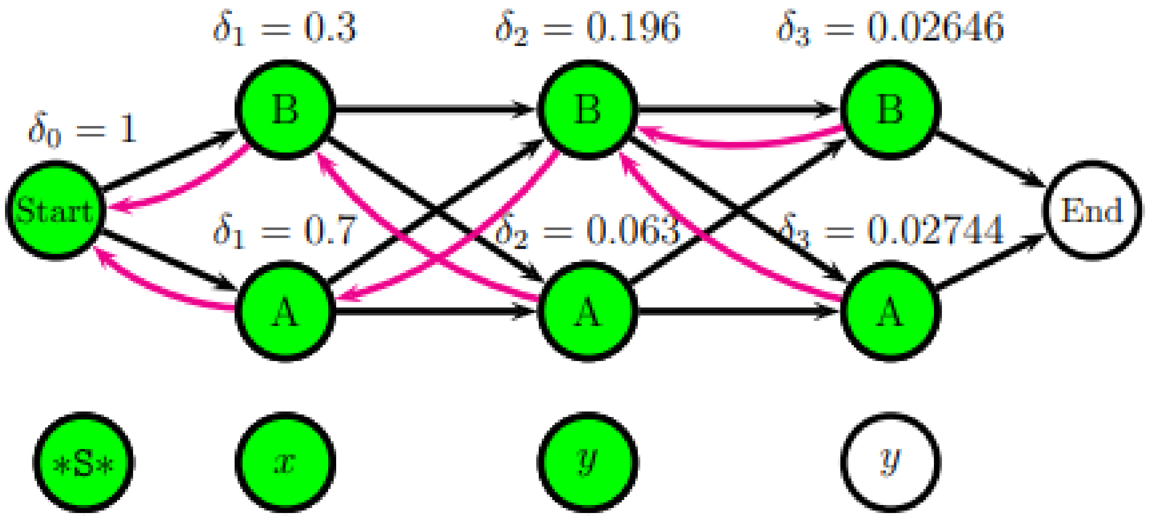
	Word		
State	*S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

Hidden Markov 概念

$$\delta_3(A) = max\{0.2 \times 0.6 \times 0.063, 0.7 \times 0.7 \times 0.196\}$$

$$\psi_1(A) = B_1 \qquad \psi_1(B) = A_1$$

$$\delta_2(B) = max\{0.7 \times 0.6 \times 0.063, 0.2 \times 0.7 \times 0.196\}$$



	Next		
Current	A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

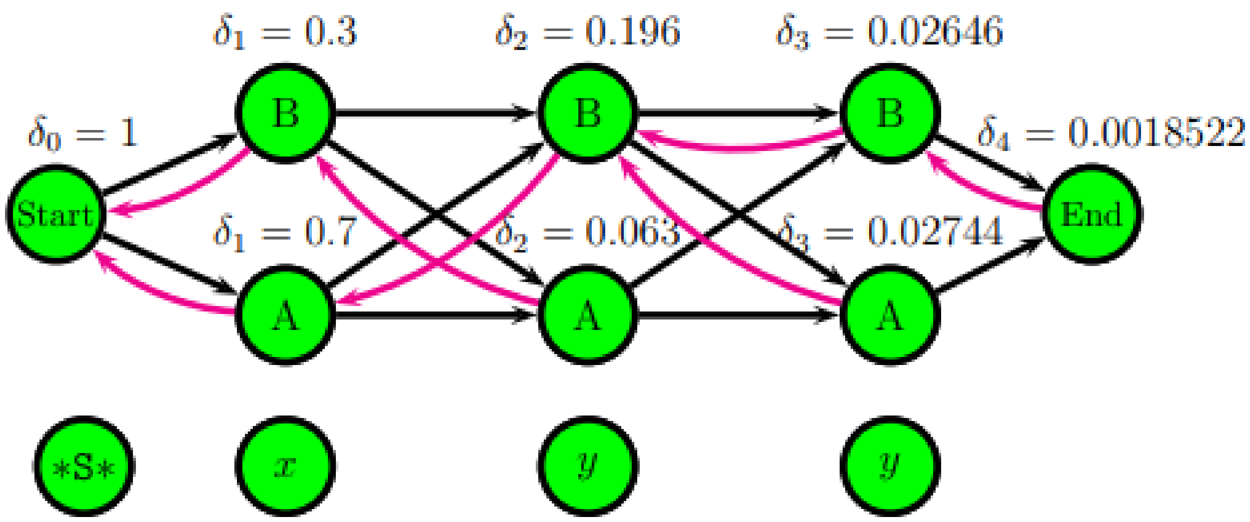
	Word		
State	*S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

Hidden Markov 概念

$$\delta_4(End) = max\{0.1 \times 0.6 \times 0.02744, 0.1 \times 0.7 \times 0.2646\}$$

$$\psi_4(End) = ABB$$

$$P(ABB, xyy) = 0.00185$$



	Next		
Current	A	B	End
Start	0.7	0.3	0
A	0.2	0.7	0.1
B	0.7	0.2	0.1

	Word		
State	*S*	x	y
Start	1	0	0
A	0	0.4	0.6
B	0	0.3	0.7

Hidden Markov 概念