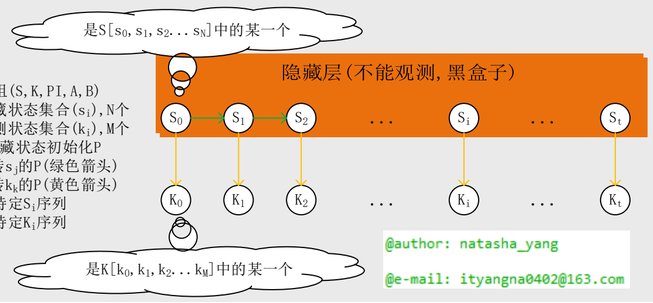
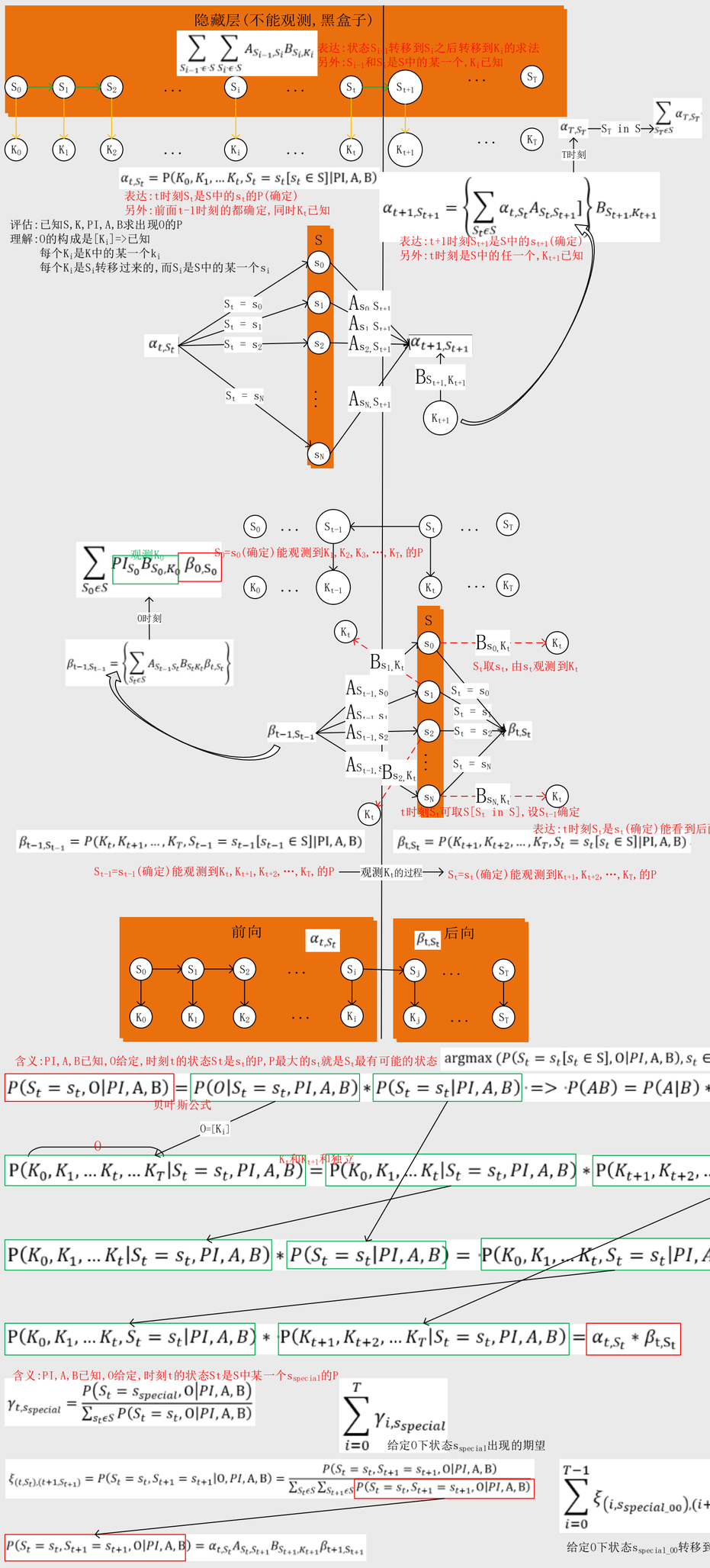
基础概念



问题一:评估(给定O,求此O出现的P)



$$\alpha_{t,S_t} = P(K_0, K_1, \dots K_t, S_t = s_t[s_t \in S] | PI, A, B)$$

$$\alpha_{t+1,S_{t+1}} = \left\{\sum_{S_t \in S} \alpha_{t,S_t} A_{S_t,S_{t+1}}\right\} B_{S_{t+1},K_{t+1}}$$

$$\beta_{t-1,S_{t-1}} = P(K_t, K_{t+1}, \dots, K_T, S_{t-1} = s_{t-1}[s_{t-1} \in S] | PI, A, B)$$

$$\beta_{t,S_t} = P(K_{t+1}, K_{t+2}, \dots, K_T, S_t = s_t[s_t \in S] | PI, A, B)$$

$$\beta_{t-1,S_{t-1}} = \left\{\sum_{S_t \in S} A_{S_{t-1},S_t} B_{S_t,K_t} \beta_{t,S_t}\right\}$$

$$P(S_t = s_t, O|PI, A, B) = P(O|S_t = s_t, PI, A, B) * P(S_t = s_t|PI, A, B) \Rightarrow P(AB) = P(A|B) * P(B)$$

$$P(K_0, K_1, \dots K_t, \dots K_T | S_t = s_t, PI, A, B) = P(K_0, K_1, \dots K_t | S_t = s_t, PI, A, B) * P(K_{t+1}, K_{t+2}, \dots K_T | S_t = s_t, PI, A, B)$$

$$P(K_0, K_1, \dots K_t | S_t = s_t, PI, A, B) * P(S_t = s_t | PI, A, B) = P(K_0, K_1, \dots K_t, S_t = s_t | PI, A, B)$$

$$P(K_0, K_1, \dots K_t, S_t = s_t | PI, A, B) * P(K_{t+1}, K_{t+2}, \dots K_T | S_t = s_t, PI, A, B) = \alpha_{t,S_t} * \beta_{t,S_t}$$

$$\gamma_{t,s_{special}} = \frac{P(S_t = s_{special}, O|PI, A, B)}{\sum_{s_t \in S} P(S_t = s_t, O|PI, A, B)}$$

$$\sum_{i=0}^{T} \gamma_{i,s_{special}}$$

$$\xi_{(t,S_t),(t+1,S_{t+1})} = P(S_t = s_t, S_{t+1} = s_{t+1}|O, PI, A, B) = \frac{P(S_t = s_t, S_{t+1} = s_{t+1}, O|PI, A, B)}{\sum_{S_t \in S} \sum_{S_{t+1} \in S} P(S_t = s_t, S_{t+1} = s_{t+1}, O|PI, A, B)}$$

$$P(S_t = s_t, S_{t+1} = s_{t+1}, O|PI, A, B) = \alpha_{t,S_t} A_{S_t,S_{t+1}} B_{S_{t+1},K_{t+1}} \beta_{t+1,S_{t+1}}$$

$$\sum_{i=0}^{T-1} \xi_{(i,s_{special\_00}),(i+1,s_{special\_01})}$$

问题二:预测(给定O,求出最可能出现的X)解码

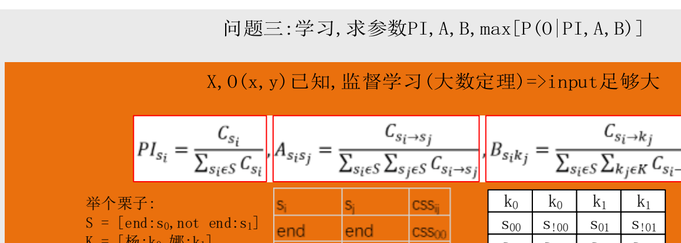
预测:已知S,K,PI,A,B,O指定,使得O出现的最有可能的X
解释:问题一=>在O下St在S中st最有可能(近似,单一求)
没有考虑某些状态之间转移P是0或很低
整体最大不能保证
解法:vitebi算法,动态规划求最短路径

t时刻St所有单个路径中P最大值

$$\delta_{t,S_t} = max_{S_0,S_1,\dots,S_{t-1}} P(S_t = s_t, S_{t-1}, \dots, S_0, K_t, \dots K_0|\theta)$$

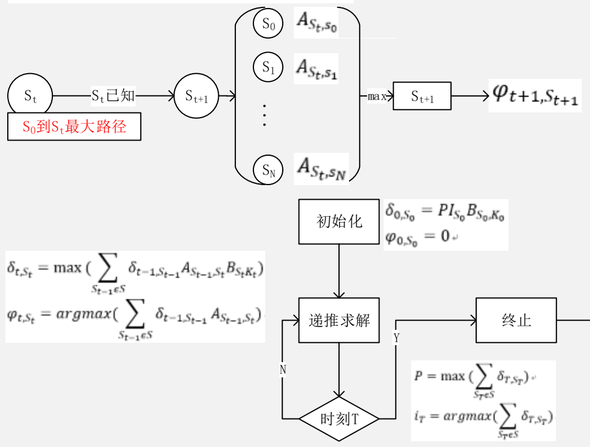
$$\delta_{t+1,S_{t+1}} = max_{S_0,S_1,\dots,S_t} P(S_{t+1} = s_{t+1}, S_t, \dots, S_0, K_{t+1}, \dots K_0|\theta) = \max\left(\sum_{S_t \in S} \delta_{t,S_t} A_{S_t,S_{t+1}} B_{S_{t+1}K_{t+1}}\right)$$

St到St+1最大路径



t时刻所有单个路径中P最大路径的第t-1个节点

$$\varphi_{t,S_t} = argmax\left(\sum_{S_{t-1} \in S} \delta_{t-1,S_{t-1}} A_{S_{t-1},S_t}\right)$$



$$\delta_{0,S_0} = PI_{S_0} B_{S_0,K_0}$$

$$\varphi_{0,S_0} = 0$$

$$\delta_{t,S_t} = \max\left(\sum_{S_{t-1} \in S} \delta_{t-1,S_{t-1}} A_{S_{t-1},S_t} B_{S_t,K_t}\right)$$

$$\varphi_{t,S_t} = argmax\left(\sum_{S_{t-1} \in S} \delta_{t-1,S_{t-1}} A_{S_{t-1},S_t}\right)$$

$$P = \max\left(\sum_{S_T \in S} \delta_{T,S_T}\right)$$

$$i_T = argmax\left(\sum_{S_T \in S} \delta_{T,S_T}\right)$$

问题三:学习,求参数PI,A,B,max[P(O|PI,A,B)]

X,O(x,y)已知,监督学习(大数定理)=>input足够大

$$PI_{s_i} = \frac{C_{s_i}}{\sum_{s_i \in S} C_{s_i}}, A_{s_i s_j} = \frac{C_{s_i \to s_j}}{\sum_{s_i \in S} \sum_{s_j \in S} C_{s_i \to s_j}}, B_{s_i k_j} = \frac{C_{s_i \to k_j}}{\sum_{s_i \in S} \sum_{k_j \in K} C_{s_i \to k_j}}$$

举个栗子:
S = [end:s0,not end:s1]
K = [杨:k0,娜:k1]

语料中第一个字是终止字(end)的个数cs0,不是终止字(not end)cs1
PI0=cs0/(cs0+cs1)
PI1=cs1/(cs0+cs1)

学习:已知X,O求参数,监督学习
已知O,求参数,无监督学习
解法:利用大数定理估计
Baum-Welch算法(EM算法)

s_i	s_j	css_{ij}
end	end	css_{00}
end	not end	css_{01}
not end	end	css_{10}
not end	not end	css_{11}

k_0	k_0	k_1	k_1
s_{00}	$s_{!00}$	s_{01}	$s_{!01}$
s_{10}	$s_{!10}$	s_{11}	$s_{!11}$

$A_{0,0}=css_{00}/(css_{00}+css_{01}+css_{10}+css_{11})$
$A_{0,1}=css_{01}/(css_{00}+css_{01}+css_{10}+css_{11})$
$A_{1,0}=css_{10}/(css_{00}+css_{01}+css_{10}+css_{11})$
$A_{1,1}=css_{11}/(css_{00}+css_{01}+css_{10}+css_{11})$
css_{ij}表示第t个字是状态i,t+1个字是状态j的个数

cs_{00}表示k_0[杨]是s_0[end]的个数,$cs_{!00}$表示k_0[杨]不是s_0[end]的个数
$B_{0,[k]0}=cs_{00}/(cs_{00}+cs_{!00})$
$B_{1,[k]0}=cs_{10}/(cs_{10}+cs_{!10})$
$B_{0,[k]1}=cs_{01}/(cs_{01}+cs_{!01})$
$B_{1,[k]1}=cs_{11}/(cs_{11}+cs_{!11})$

O为观测数据,X为隐数据

1.确实完全数据的对数似然函数

O=[Ki],X=[Si]=>(O,X)=[Ki,Si]
L(θ)=log(P(O,X|θ))

$$L(\theta) = \sum_{i=0}^{T} \log P(K_i|\theta) = \sum_{i=0}^{T} \sum_{j=0}^{T} \log P(K_i, S_j|\theta)$$

2.step E,求Q函数

在分布j下的对数似然函数
在当前的θ下属于分布j的概率
参数的当前估计值
与隐状态无关,是常量可省略
Q函数

$$Q(\theta, \bar{\theta}) = \sum_{j=0}^{T} \log P(O, S_j|\theta) P(S_j|O, \bar{\theta}) = \sum_{j=0}^{T} \log P(O, S_j|\theta) \frac{P(O, S_j|\theta)}{P(O|\theta)} \propto \sum_{j=0}^{T} \log P(O, S_j|\theta) P(O, S_j|\theta)$$

$$\sum_{j=0}^{T} \log P(O, S_j|\theta) P(O, S_j|\bar{\theta}) = \left\{\sum_{S_0 \in S} \log PI_{S_0} + \sum_{j=0}^{T-1} \sum_{S_j S_{j+1} \in S} \log(A_{S_j,S_{j+1}}) + \sum_{j=0}^{T} \sum_{S_j \in S} \log(B_{S_j,K_j})\right\} P(O, S_j|\bar{\theta})$$

$$P(O|\theta) = \sum_{j=0}^{T} P(O, S_j|\theta) = \sum_{j=0}^{T} P(O|S_j, \theta) P(S_j|\theta) = \sum_{S_0 \in S} PI_{S_0} B_{S_0,K_0} \sum_{j=0}^{T-1} \sum_{S_j S_{j+1} \in S} A_{S_j,S_{j+1}} B_{S_{j+1},K_{j+1}}$$

3.step M,极大化Q函数

求PI,拉格朗日函数
约束:加和为1

$$\sum_{S_0 \in S} \log PI_{S_0} P(O, S_0|\bar{\theta}) + \gamma\left(\sum_{S_0 \in S} PI_{S_0} - 1\right)$$

求A,拉格朗日函数
约束:加和为1
固定状态转到下一个Sx就是A的某一行

$$\sum_{j=0}^{T-1} \sum_{S_j S_{j+1} \in S} \log(A_{S_j,S_{j+1}}) P(O, S_j, S_{j+1}|\bar{\theta}) + \gamma\left(\sum_{S_j \in S} A_{S_i,S_j} - 1\right)$$

求B,拉格朗日函数
约束:加和为1
固定状态S_x转移到每个可观测状态的P

$$\sum_{j=0}^{T} \sum_{S_j \in S} \log(B_{S_j,K_j}) P(O, S_j|\bar{\theta}) + \gamma\left(\sum_{K_j \in K} B_{S_j,K_j} - 1\right)$$