第10章，Hidden Markov Model (HMM)

背景：

概率图
- 有向 — Bayesian Network
- 无向 — Markov Random Field
  (Markov Network)

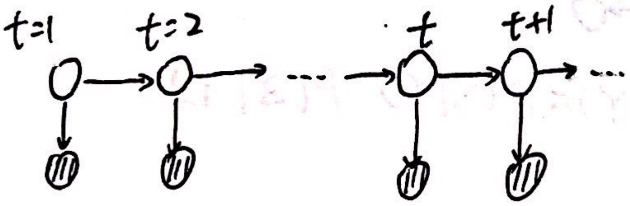
+time

→ Dynamic Model
- HMM — 离散
- Kalman Filter — 连续线性
- Particle Filter — 连续非线性

state

Xi之间不是iid
- time
- mixture

HMM



第10章 隐马尔可夫模型

背景 隐马尔可夫模型的基本概念

变量多,用概率图模型(有向图)表示变量之间关系

模型参数及符号:

状态集合

观测集合

初始状态概率向量

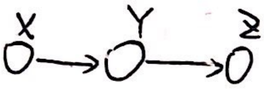
表示Y受到X影响

表示看的见的变量

表示隐变量

★ 变量可见与否会影响独立性.例

①



$P(X,Z) \neq p(x)\cdot P(z)$

$P(X,Z|Y)=P(X|Y)\cdot P(Z|Y)$

②

$P(X,Z) \neq P(X)P(Z)$

$P(X,Z|Y)= P(X|Y)\cdot P(Z|Y)$

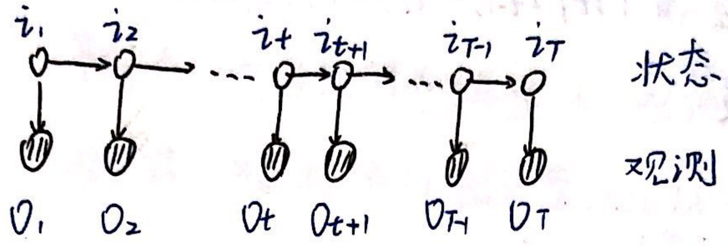
(如不知Y值,需从Z推Y,再用Y推X.会有影响

③

$P(X,Z)=P(X)\cdot P(Z)$

$P(X,Z|Y) \neq P(Z|Y)\cdot P(X|Y)$

HMM概率图模型



$i_t \in \{q_1, \cdots q_N\}$　　$I = \{i_1, \cdots, i_T\}$

$o_t \in \{v_1, \cdots, v_M\}$　　$O = \{o_1, \cdots, o_T\}$

①

$$\begin{array}{c|cccc} & i_2=a_1 & i_2=q_2 & \cdots & i_2=q_N \\ \hline i_1=q_1 & a_{11} & a_{12} & \cdots & a_{1N} \\ i_1=q_2 & a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & & \vdots & & \\ i_1=q_N & a_{N1} & a_{N2} & \cdots & a_{NN} \end{array}$$

$(a_{ij}=P(i_2=q_j \mid i_1=q_1))$

→ 和为1（∵ a_{ij}为条件概率）

→ $A_{N\times N}$（状态之间有向边）

状态转移关系矩阵

★（不随时间变化）→ i_t　i_{t+1}

②

$$\begin{array}{c|cccc} & o_1=v_1 & o=v_2 & \cdots & o_1=v_M \\ \hline i_1=q_1 & b_1(1) & b_1(2) & \cdots & b_1(M) \\ i_2=q_2 & b_2(1) & b_2(2) & \cdots & b_2(M) \\ \vdots & & \vdots & & \\ i_1=q_N & b_N(1) & b_N(2) & \cdots & b_N(M) \end{array}$$

$b_j(k)=P(o_t=v_k \mid i_t=q_j)$

和为1

（在时刻t处于状态q_j的条件下生成观测v_k的概率）

$B_{N\times M}$

观测概率矩阵

i_1　o_1

③

$$\begin{bmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_N \end{bmatrix} \begin{matrix} =P(i_1=q_1) \\ =P(i_1=q_2) \\ \vdots \\ =P(i_1=q_N) \end{matrix}$$

π（初始状态）

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

N — $\boxed{N-1}$ → 自由参数个数

$N\times N$ — $\boxed{N\times N-N}$

$N\times M$ — $\boxed{N\times M-N}$

$N-1$表示减去$(\pi_1+\pi_2\cdots+\pi_N)$概率和为1这个约束

10.1

两个基本假设：（为了简化模型中变量之间的关系）

齐次马尔科夫性 $P(i_t|i_{t-1},\cdots i_1)=P(i_t|i_{t-1})$

观测独立假设

三个基本问题（对应3小节）

概率计算问题 $P(O|\lambda)$

学习问题 $\underset{\lambda}{\arg\max}\, P(O|\lambda)$

预测问题 $P(I|O)$ 可看成让计算机自动标注词性

10.2 概率计算法

计算 $P(O|\lambda)$

1.直接计算法

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

$$=\sum_{i_1,i_2,\cdots i_T}\pi_{i_1}b_{i_1}(o_1)a_{i_1i_2}b_{i_2}(o_2)\cdots a_{i_{T-1}i_T}b_{i_T}(o_T)$$

计算复杂度：$O(TN^T)$

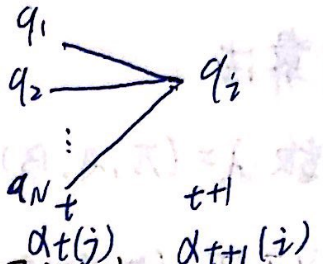
$i_1\rightarrow N$

$i_2\rightarrow N$

$\vdots$

$i_T\xrightarrow{i}N$

2 前向算法



计算复杂度 $O(TN^2)$

输入：隐马尔可夫模型λ，观测序列O.

输出：观测序列概率 $P(O|\lambda)$

(1) 初值

$$\alpha_1(i) = \pi_i b_i(o_1)。\quad i=1,2,\cdots,N$$

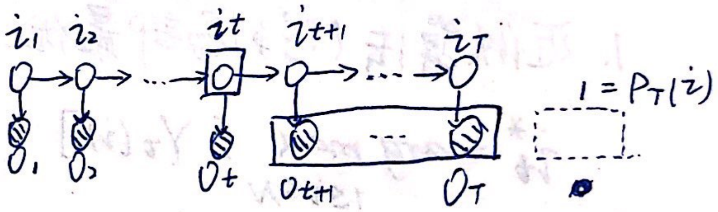
(2) 递推　对 $t=1,2,\cdots,T-1$,

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

(3) 终止

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

3 后向算法



计算复杂度 $O(TN^2)$

输入：隐马尔可夫模型λ，观测序列O.

输出：观测序列概率 $P(O|\lambda)$

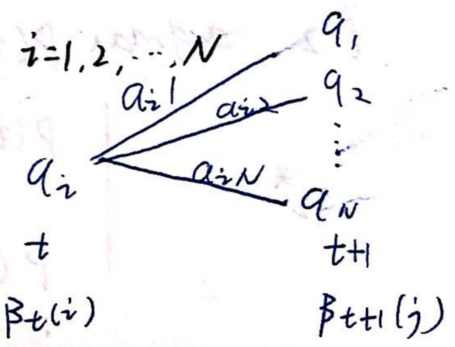
$$\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T | i_t = q_i, \lambda)$$

(1) $\beta_T(i)=1,\quad i=1,2,\cdots N$

(2) 对 $t=T-1, T-2, \cdots, 1$

$$\beta_t(i) = \sum_{j=1}^{N} a_{ij} b_j(o_{t+1}) \beta_{t+1}(j), \quad i=1,2,\cdots,N$$

(3) $$P(O|\lambda) = \sum_{i=1}^{N} \pi_i b_i(o_1) \beta_1(i)$$



10.3 学习算法

估计参数 $\lambda=(\pi, A, B)$

1. 监督学习方法（知道O与I） P_{204}

2. Buam-welch算法（EM算法）

10.4 预测算法

目标：$\arg\max P(I|O,\lambda)$

1. 近似算法（得到局部最优，没考虑状态之间联系）

$$i_t^* = \arg\max_{1\le i\le N} [\gamma_t(i)], \quad t=1,2,\cdots,T$$

$$\gamma_t(i) = P(i_t = q_i|O,\lambda)$$

步骤

$t=1 \quad \arg\max_{i} P(i_1=q_i|O,\lambda)$

$$i_1^* \quad \left| \begin{array}{l} P(i_1=q_1|O,\lambda) \\ \vdots \\ P(i_1=q_N|O,\lambda) \end{array} \right.$$

$$I^*=(i_1^*, \cdots, i_T^*)$$

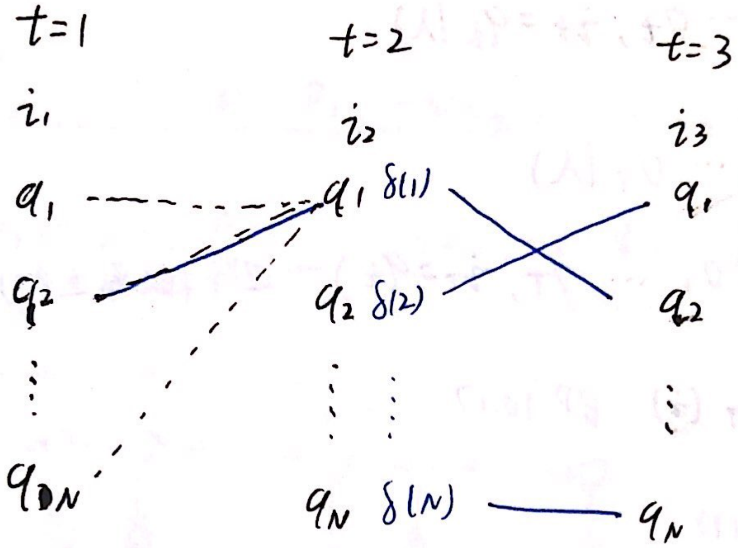
$t=2. \quad \arg\max P(i_2=q_i|O,\lambda)$

$$i_2^* \quad \left| \begin{array}{l} P(i_2=q_1|O,\lambda) \\ \vdots \\ P(i_2=q_N|O,\lambda) \end{array} \right.$$

$\vdots$

2. 维特比算法（全局最优）

例



相当于用动态规划求概率最大路径（一个路径对应一个状态序列）

$$\delta_t(i) = \underset{i_{t-1},\cdots,i_1}{\operatorname{argmax}} P(\underline{i_t = q_i}, i_{t-1}, \cdots i_1, O_t, O_{t-1}, \cdots O_1 | \lambda)$$

$$\because P(I,O) = \underbrace{P(O)}_{\text{一直不变}} \cdot P(O|I)$$

∴最大化 $P(I,O)$ 相当于 $P(O|I)$

前向算法 2个推导 $\left\{ \begin{array}{l} \text{10.16 递推公式} \\ \text{10.17 终止公式} \end{array} \right.$

$$\alpha_t(i) = P(o_1, \cdots o_t, i_t = q_i \mid \lambda)$$

目标是要求 $P(o_1, \cdots, o_T \mid \lambda)$

$$= \sum_{i=1}^{N} P(o_1, \cdots, o_T, i_T = q_i) \to \text{(边际概率之和)}$$

$$= \sum_{i=1}^{N} \alpha_T(i) \quad \text{即 10.17}$$

即

$$\begin{array}{ccccc} \alpha_1(1) & & \alpha_2(1) & & \alpha_T(1) \\ \vdots & \to & \vdots & \to \cdots & \vdots \\ \alpha_1(N) & & \alpha_2(N) & & \alpha_T(N) \end{array}$$

找两步之间的关系

$$\alpha_t(j) = P(o_1, \cdots, o_t, \underline{i_t} = q_j \mid \lambda)$$

$$\alpha_{t+1}(i) = P(o_1, \cdots, o_t, \underline{o_{t+1}}, \underline{i_{t+1}} = q_i \mid \lambda) \quad \text{——} ①$$

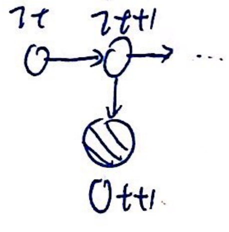
②是①的边际概率

$$= \sum_{j=1}^{N} P(o_1, \cdots, o_t, o_{t+1}, i_{t+1} = q_i, i_t = q_j) \text{ — } ②$$

$$= \sum_{j=1}^{N} \underbrace{P(o_1, \cdots, o_t, i_t = q_j)}_{\alpha_t(j)} \cdot \overset{\to b_i(o_{t+1})}{\underline{P(o_{t+1} \mid i_{t+1} = q_i)}} \cdot \underbrace{P(i_{t+1} = q_i \mid i_t = q_j)}_{a_{ji}}$$

$$= \left[\sum_{j=1}^{N} \alpha_t(j) a_{ji} \right] b_i(o_{t+1})$$

即 10.16



维特比算法

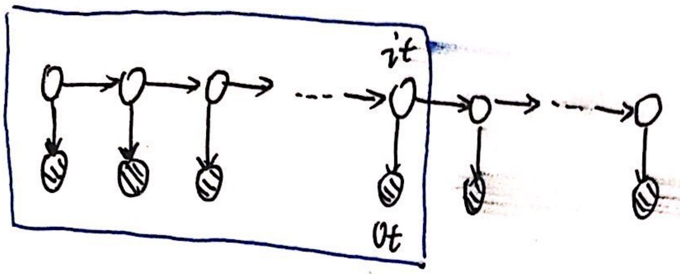
要:

$$\max P(i_1,\cdots i_T \mid O_1,\cdots,O_T)$$

$$\overset{\max}{\Rightarrow} \boxed{P(i_1,\cdots,i_T,O_1,\cdots O_T)} \cdot \underline{P(O_1,\cdots O_T)} \hookrightarrow 不变$$

$$\delta_t(i) = \max_{i_1,\cdots i_{t-1}} P(\underbrace{i_1,\cdots,i_{t-1}}_{无约束}, \underbrace{i_t=i}_{约束}, O_t,\cdots,O_1)$$

$\|$



$$\delta_{t+1}(i) = \arg\max_{i_1,\cdots,i_t} P(i_1,\cdots,i_t, \underline{i_{t+1}=i}, \underline{O_{t+1}},\cdots,O_1)$$

★ 贝尔曼最优原理 (只看 it, it 之间的路径不会变)

$$= \arg\max_j \ \delta_t(j)\cdot P(\underset{\substack{\downarrow \\ i_{t+1}\text{只和}i_t\text{有关} \\ \text{即 } \underline{a_{ji}}}}{i_{t+1}\mid i_t=j})\cdot P(\underset{\substack{\downarrow \\ O_{t+1}\text{只和}i_{t+1}\text{有关} \\ \text{即 } \underline{b_i(O_{t+1})}}}{O_{t+1}\mid i_{t+1}=i})$$

$$= \underbrace{\max_j [\delta_t(j)\, a_{ji}]}\ \underline{b_i(O_{t+1})} \hookrightarrow 不变$$

对于每个时刻,都需记录N个不同取值的路径

∵之后随着状态改变, 当前概率最大的路径会改变