第十一章

条件随机场

袁春 清华大学深圳研究生院
李航 华为诺亚方舟实验室

目录

一、概率无向图模型

- 概念：
  - 概率无向图模型(probabilistic undirected graphical model)
  - 马尔可夫随机场(Markov random field)
  - 可以由无向图表示的联合概率分布。

模型定义

- Graph
- Node
- Edge
- v，集合V
- e，集合E
- G=（V，E）
- 结点v，随机变量Y_v；边e，随机变量间的概率依赖关系
- 概率图模型(Probabilistic graphical model): 用图表示的概率分布。

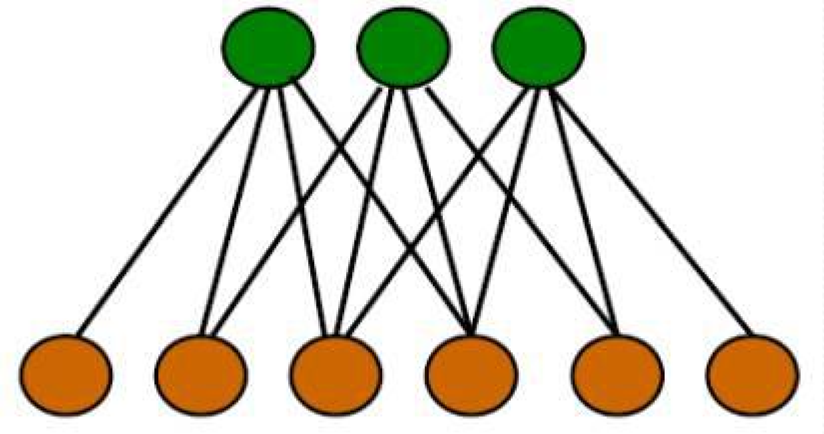
模型定义

- 定义：
- 给定一个联合概率分布P(Y)和表示它的无向图G，
- 定义无向图表示的随机变量之间存在的
  - 成对马尔可夫性(pairwise Markov property)
  - 局部马尔可夫性(local Markov properly)
  - 全局马尔可夫性(global Markov property)

模型定义

- 成对马尔可夫性(Pairwise Markov property)
  - 设u和v是无向图G中任意两个没有边连接的结点，结点u和v分别对应随机变量Yu和Yv，
  - 其他所有结点为O，对应的随机变量组是Yo
  - 给定随机变量组Yo的条件下随机变量Yu和Yv是条件独立的

$$P(Y_u, Y_v \mid Y_O) = P(Y_u \mid Y_O) P(Y_v \mid Y_O)$$



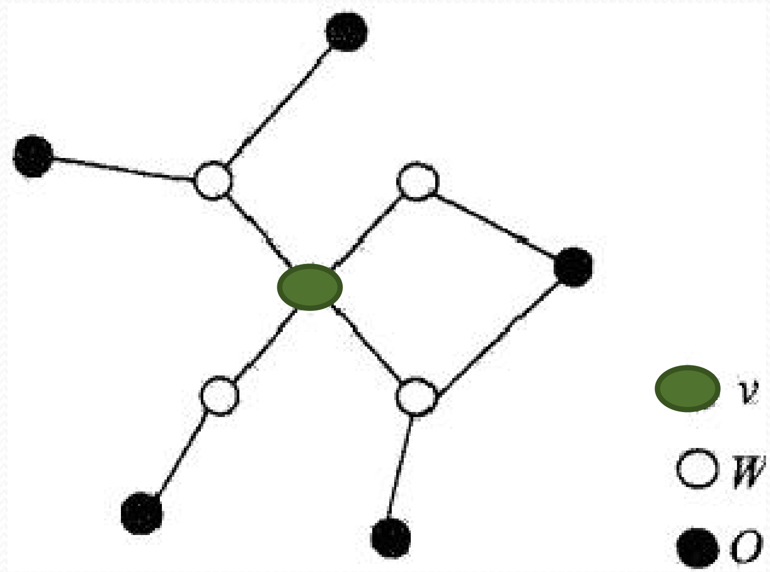
模型定义

- 局部马尔可夫性(Local Markov properly)
  - v 任意结点
  - W与v有边相连
  - O 其它

$$P(Y_v, Y_O \mid Y_W) = P(Y_v \mid Y_W) P(Y_O \mid Y_W)$$

- 在 $P(Y_O \mid Y_W) > 0$ 时，等价于

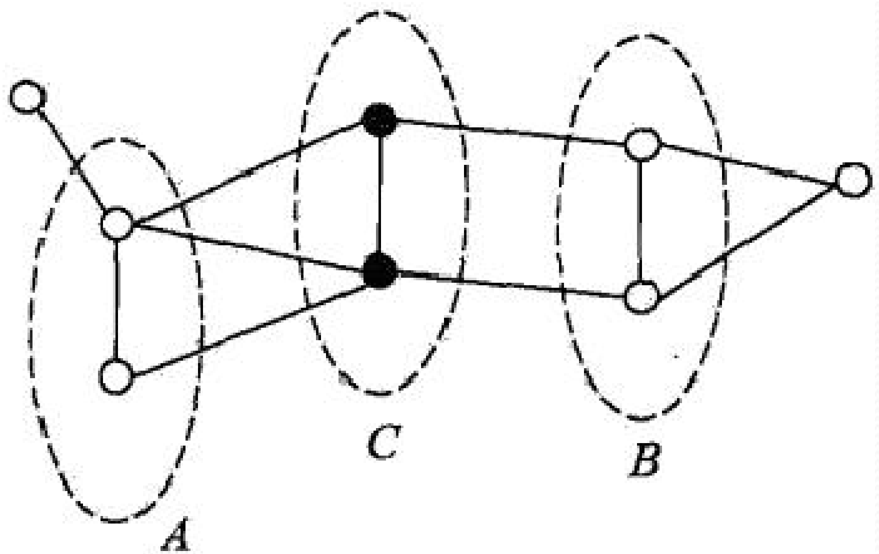
$$P(Y_v \mid Y_W) = P(Y_v \mid Y_W, Y_O)$$



模型定义

- 全局马尔可夫性(Global Markov property)
  - 结点集合A, B是在无向图G中被结点集合C分开的任意结点集合，

$$P(Y_A, Y_B \mid Y_C) = P(Y_A \mid Y_C) P(Y_B \mid Y_C)$$



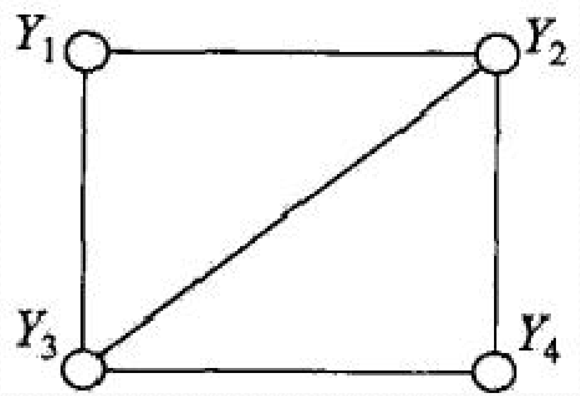
模型定义

- 概率无向图模型：
  - 设有联合概率分布P(Y),由无向图G=(V, E)表示，在图G中，结点表示随机变量，边表示随机变量之间的依赖关系,
  - 如果联合概率分布P(Y)满足成对、局部或全局马尔可夫性，就称此联合概率分布为概率无向图模型(probability undirected graphical model)，或马尔可夫随机场(Markov random field).
  - 问题关键：求联合概率，引申为对联合概率进行因子分解。

概率无向图模型的因子分解

- 定义：团、最大团
- 无向图G中任何两个结点均有边连接的结点子集称为团(clique)。
- 若C是无向图G的一个团，并且不能再加进任何一个c的结点使其成为一个更大的团，则称此C为最大团(maximal clique).

- 两个结点的团？
- 三个结点的团？



概率无向图模型的因子分解

- 将概率无向图模型的联合概率分布表示为其最大团上的随机变量的函数的乘积形式的操作，称为概率无向图模型的因子分解(Factorization).
- 给定概率无向图模型，设其无向图为G，C为G上的最大团，Yc表示C对应的随机变量，那么概率无向图模型的联合概率分布P(Y)可写作图中所有最大团C上的函数 $\Psi_C(Y_C)$ 的乘积形式，即

$$P(Y)=\frac{1}{Z}\prod_{C}\Psi_C(Y_C)$$

- Z是规范化因子(normalization factor)

$$Z=\sum_{Y}\prod_{C}\Psi_C(Y_C)$$

概率无向图模型的因子分解

- 势函数：

$$\Psi_C(Y_C)=\exp\{-E(Y_C)\}$$

- 定理11.1 (Hammersley-Clifford定理)：概率无向图模型的联合概率分布P(Y)可以表示为如下形式:

$$P(Y)=\frac{1}{Z}\prod_C \Psi_C(Y_C)$$

$$Z=\sum_Y\prod_C \Psi_C(Y_C)$$

二、条件随机场的定义与形式

- 条件随机场(conditional random field)的定义：
  - 给定随机变量X条件下，随机变量Y的马尔可夫随机场。
- 定义在线性链上的特殊的条件随机场：
  - 线性链条件随机场(linear chain conditional random field)
  - 线性链条件随机场可以用于标注等问题；
  - 在条件概率模型P(Y|X)中，Y是输出变量，表示标记序列，X是输入变量，表示需要标注的观测序列，也把标记序列称为状态序列。

条件随机场的定义与形式

- 条件随机场(conditional random field)三个主要问题：
- 概率计算
- 模型学习
- 推测状态

条件随机场的定义与形式

- 条件随机场：
  - 设X与Y是随机变量，P(Y|X)是在给定X的条件下Y的条件概率分布，若随机变量Y构成一个由无向图G=(V,E)表示的马尔可夫随机场，即满足马尔科夫性

$$P(Y_v \mid X, Y_w, w \neq v) = P(Y_v \mid X, Y_w, w \sim v)$$

  - 对任意结点v成立，则称条件概率分布P(Y|X)为条件随机场，式中w~v表示在图G=(V,E)中与结点v有边连接的所有结点w，w≠v表示结点v以外的所有结点。

条件随机场的定义与形式

- 线性链情况：

$$G = (V = \{1, 2, \cdots, n\},\ E = \{(i, i+1)\})\text{，}\ i = 1, 2, \cdots, n-1$$

$$X = (X_1, X_2, \cdots, X_n)\text{，}\ Y = (Y_1, Y_2, \cdots, Y_n)$$

- 最大团是相邻两个结点的集合, 线性链条件随机场:

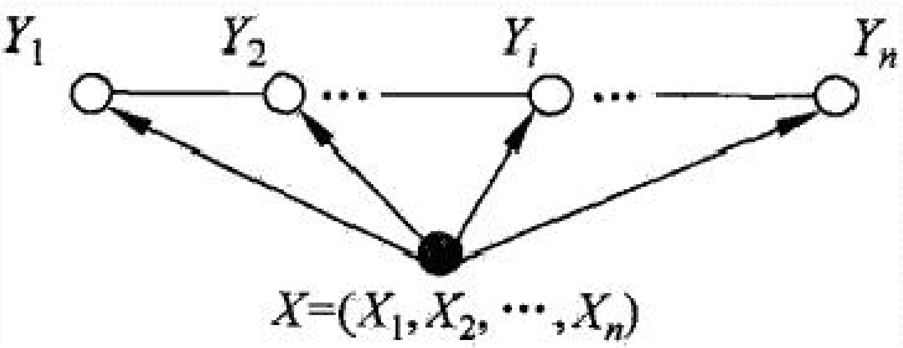
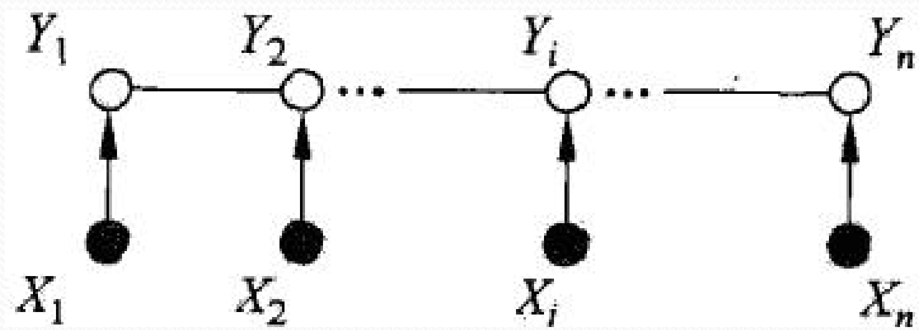


图 11.4　线性链条件随机场



X 和 Y 有相同的图结构的线性链条件随机场

条件随机场的定义与形式

- 定义(线性链条件随机场)
- 设 $X=(X_1,X_2,\cdots,X_n)$，$Y=(Y_1,Y_2,\cdots,Y_n)$ 均为线性链表示的随机变量序列，若在给定随机变量序列X的条件下，随机变量序列Y的条件概率分布P(Y|X)构成条件随机场。即满足马尔可夫性

$$P(Y_i \mid X,Y_1,\cdots,Y_{i-1},Y_{i+1},\cdots,Y_n)=P(Y_i \mid X,Y_{i-1},Y_{i+1})$$

$i=1,2,\cdots,n$（在 $i=1$ 和 n 时只考虑单边）

- 则称P(Y |X)为线性链条件随机场。
- 在标注问题中，X表示输入观测序列，Y表示对应的输出标记序列或状态序列.

条件随机场的参数化形式

- 定理:
- (线性链条件随机场的参数化形式): 设P(Y|X)为线性链条件随机场，则在随机变量X取值为x的条件下，随机变量Y取值为y的条件概率具有如下形式:

$$P(y|x)=\frac{1}{Z(x)}\exp\left(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\right)$$

- 其中: $$Z(x)=\sum_y \exp\left(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,l}\mu_l s_l(y_i,x,i)\right)$$
- t_k 定义在边上的特征函数，转移特征，依赖于前一个和当前位置，
- s_l 定义在结点上的特征函数，状态特征，依赖于当前位置

条件随机场的参数化形式

- 例：标准问题，输入观测为：X=(X1,X2,X3), 输出标记为
- Y=(Y1,Y2,Y3), Y1,Y2,Y3 取值于{1,2}
- 假设特征和对应权值，只注明特征取值为1，为0省略

$$t_1 = t_1(y_{i-1}=1, y_i=2, x, i)\text{，}\quad i=2,3\text{，}\quad \lambda_1 = 1$$

$$t_1(y_{i-1}, y_i, x, i) = \begin{cases} 1, & y_{i-1}=1, y_i=2, x, i, (i=2,3) \\ 0, & \text{其他} \end{cases}$$

$$t_2 = t_2(y_1=1, y_2=1, x, 2)\quad \lambda_2 = 0.5$$

$$t_3 = t_3(y_2=2, y_3=1, x, 3)\quad \lambda_3 = 1$$

$$t_4 = t_4(y_1=2, y_2=1, x, 2)\quad \lambda_4 = 1$$

$$t_5 = t_5(y_2=2, y_3=2, x, 3)\quad \lambda_5 = 0.2$$

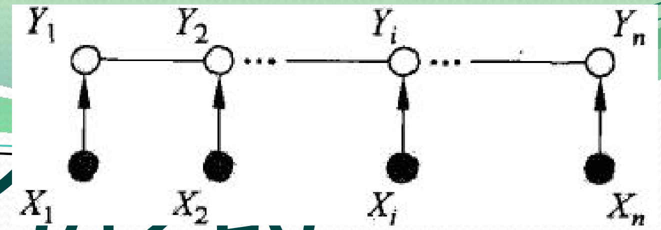
$$s_1 = s_1(y_1=1, x, 1)\text{，}\quad \mu_1 = 1$$

$$s_2 = s_2(y_i=2, x, i)\text{，}\quad i=1,2\quad \mu_2 = 0.5$$

$$s_3 = s_3(y_i=1, x, i)\text{，}\quad i=2,3\quad \mu_3 = 0.8$$

$$s_4 = s_4(y_3=2, x, 3)\text{，}\quad \mu_4 = 0.5$$

条件随机场的参数化形式



$$P(y\mid x)\propto \exp\left[\sum_{k=1}^{5}\lambda_k\sum_{i=2}^{3}t_k(y_{i-1},y_i,x,i)+\sum_{k=1}^{4}\mu_k\sum_{i=1}^{3}s_k(y_i,x,i)\right]$$

- 对给定的观测序列x，标记序列Y=(1, 2, 2)的非规范化条件概率为

$$P(y_1=1,y_2=2,y_3=2\mid x)\propto \exp(\quad)$$

$$t_1=t_1(y_{i-1}=1,y_i=2,x,i),\quad i=2,3,\quad \lambda_1=1$$

$$t_2=t_2(y_1=1,y_2=1,x,2)\quad \lambda_2=0.5$$
$$t_3=t_3(y_2=2,y_3=1,x,3)\quad \lambda_3=1$$
$$t_4=t_4(y_1=2,y_2=1,x,2)\quad \lambda_4=1$$
$$t_5=t_5(y_2=2,y_3=2,x,3)\quad \lambda_5=0.2$$

$$s_1=s_1(y_1=1,x,1),\quad \mu_1=1$$
$$s_2=s_2(y_i=2,x,i),\quad i=1,2\quad \mu_2=0.5$$
$$s_3=s_3(y_i=1,x,i),\quad i=2,3\quad \mu_3=0.8$$
$$s_4=s_4(y_3=2,x,3),\quad \mu_4=0.5$$

条件随机场的简化形式

- 注意到条件随机场中同一特征在各个位置都有定义，可以对同一个特征在各个位置求和，将局部特征函数转化为一个全局特征函数，这样就可以将条件随机场写成权值向量和特征向量的内积形式，即条件随机场的简化形式。
- 首先将转移特征和状态特征及其权值用统一的符号表示，设有k1个转移特征，k2个状态特征，K=k1+k2,记

$$f_k(y_{i-1}, y_i, x, i) = \begin{cases} t_k(y_{i-1}, y_i, x, i), & k = 1, 2, \cdots, K_1 \\ s_l(y_i, x, i), & k = K_1 + l;\ l = 1, 2, \cdots, K_2 \end{cases}$$

条件随机场的简化形式

- 然后，对转移与状态特征在各个位置i求和，记作

$$f_k(y,x)=\sum_{i=1}^{n} f_k(y_{i-1},y_i,x,i),\quad k=1,2,\cdots,K$$

- 权值：

$$w_k=\begin{cases}\lambda_k, & k=1,2,\cdots,K_1\\ \mu_l, & k=K_1+l;l=1,2,\cdots,K_2\end{cases}$$

- 条件随机场可表示为：

$$P(y\mid x)=\frac{1}{Z(x)}\exp\sum_{k=1}^{K} w_k f_k(y,x)$$

$$Z(x)=\sum_{y}\exp\sum_{k=1}^{K} w_k f_k(y,x)$$

条件随机场的简化形式

- 若w表示权值向量：

$$w=(w_1,w_2,\cdots,w_K)^{\mathrm{T}}$$

- 以F(y,x)表示全局特征向量，即

$$F(y,x)=(f_1(y,x),f_2(y,x),\cdots,f_K(y,x))^{\mathrm{T}}$$

- 条件随机场写成内积：

$$P_w(y\mid x)=\frac{\exp\left(w\cdot F(y,x)\right)}{Z_w(x)}$$

$$Z_w(x)=\sum_y \exp\left(w\cdot F(y,x)\right)$$

条件随机场的矩阵形式

- 线性链条件随机场，引进特殊的起点和终点状态标记 Y_0= start， Y_{n+1} = stop，这时$P_w(y|x)$可以通过矩阵形式表示。
- 对观测序列x的每一个位置i=1,2,..n+1，定义一个m阶矩阵(m是标记Y_i取值的个数)

$$M_i(x)=[M_i(y_{i-1},y_i\mid x)]$$

$$M_i(y_{i-1},y_i\mid x)=\exp\left(W_i(y_{i-1},y_i\mid x)\right)$$

$$W_i(y_{i-1},y_i\mid x)=\sum_{i=1}^{K} w_k f_k(y_{i-1},y_i,x,i)$$

条件随机场的矩阵形式

- 给定观测序列x，标记序列y的非规范化概率可以通过n+1个矩阵的乘积表示：

$$\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i \mid x)$$

- 条件概率$P_w(y|x)$：

$$P_w(y \mid x) = \frac{1}{Z_w(x)} \prod_{i=1}^{n+1} M_i(y_{i-1}, y_i \mid x)$$

- $Z_w(x)$为规范化因子，是n+1个矩阵的乘积的(start, stop)元素

$$Z_w(x) = \left(M_1(x) M_2(x) \cdots M_{n+1}(x)\right)_{\text{start,stop}}$$

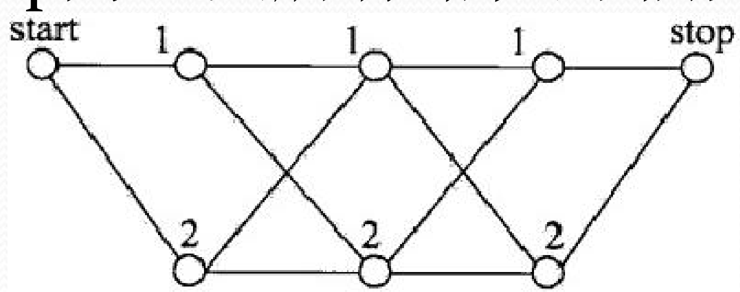
条件随机场的矩阵形式

- 例：线性链条件随机场，观测序列x，状态序列y，$i=1,2,3$ $n=3$，标记y_i属于{1，2}，假设 $y_0=start=1$，$y_4=stop=1$，各个位置的随机矩阵：

$$M_1(x)=\begin{bmatrix} a_{01} & a_{02} \\ 0 & 0 \end{bmatrix}, \quad M_2(x)=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$$

$$M_3(x)=\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}, \quad M_4(x)=\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}$$

试求状态序列y以start为起点stop为终点所有路径的非规范化概率及规范化因子。



条件随机场的矩阵形式

- 解： 首先计算从start到stop对应与y=(1,1,1), y=(1,1,2),..y=(2,2,2）各路径的非规范化概率分别是：

$$a_{01}b_{11}c_{11}, \quad a_{01}b_{11}c_{12}, \quad a_{01}b_{12}c_{21}, \quad a_{01}b_{12}c_{22}$$
$$a_{02}b_{21}c_{11}, \quad a_{02}b_{21}c_{12}, \quad a_{02}b_{22}c_{21}, \quad a_{02}b_{22}c_{22}$$

- 求规范化因子，通过计算矩阵乘积，第1行第1列的元素为：

$$\begin{aligned} &a_{01}b_{11}c_{11} + a_{02}b_{21}c_{11} + a_{01}b_{12}c_{21} + a_{02}b_{22}c_{22} \\ &\quad + a_{01}b_{11}c_{12} + a_{02}b_{21}c_{12} + a_{01}b_{12}c_{22} + a_{02}b_{22}c_{21} \end{aligned}$$

- 恰好等于从start到stop的所有路径的非规范化概率之和，及规范化因子。

三、条件随机场的概率计算问题

- 条件随机场的概率计算问题
  - 给定条件随机场P(Y|X),输入序列x和输出序列y，
  - 计算条件概率： $P(Y_i = y_i \mid x)$，$P(Y_{i-1} = y_{i-1}, Y_i = y_i \mid x)$
  - 以及相应的数学期望问题。
- 引进前向-后向向量，递归计算。

条件随机场的概率计算问题

- 前向-后向算法：
- 对每个指标i=0,1,...,n+1 , 定义前向向量 $\alpha_i(x)$
- 递推公式： $\alpha_0(y \mid x)=\begin{cases}1, & y=\text{start} \\ 0, & \text{否则}\end{cases}$

$$\alpha_i^{\mathrm{T}}(y_i \mid x)=\alpha_{i-1}^{\mathrm{T}}(y_{i-1} \mid x) M_i(y_{i-1}, y_i \mid x), \quad i=1,2,\cdots,n+1$$

- 又可表示为：

$$\alpha_i^{\mathrm{T}}(x)=\alpha_{i-1}^{\mathrm{T}}(x) M_i(x)$$

- 即表示在位置i的标记是yi，且到位置i的前部分标记序列的非规范化概率，yi可取的值m个，所以$\alpha_i(x)$是m维列向量。

条件随机场的概率计算问题

- 前向-后向算法：
- 同样，对每个指标i=0,1,...,n+1，定义后向向量$\beta_i(x)$

$$\beta_{n+1}(y_{n+1} \mid x) = \begin{cases} 1, & y_{n+1} = \text{stop} \\ 0, & \text{否则} \end{cases}$$

$$\beta_i(y_i \mid x) = M_i(y_i, y_{i+1} \mid x)\beta_{i-1}(y_{i+1} \mid x)$$

- 又可表示为： $\beta_i(x) = M_{i+1}(x)\beta_{i+1}(x)$
- 即表示在位置i的标记是yi，且从位置i+1到n的后部分标记序列的非规范化概率
- 前向-后向得： $Z(x) = \alpha_n^{T}(x) \cdot \mathbf{1} = \mathbf{1}^{T} \cdot \beta_1(x)$

条件随机场的概率计算问题

- 概率计算
- 按照前向-后向向量的定义，
- 可计算标记序列在位置i是标记y_i的条件概率
- 和在位置i-1与i是标记y_{i-1}和y_i的条件概率：

$$P(Y_i = y_i \mid x) = \frac{\alpha_i^{T}(y_i \mid x)\beta_i(y_i \mid x)}{Z(x)}$$

$$P(Y_{i-1} = y_{i-1}, Y_i = y_i \mid x) = \frac{\alpha_{i-1}^{T}(y_{i-1} \mid x)M_i(y_{i-1}, y_i \mid x)\beta_i(y_i \mid x)}{Z(x)}$$

- 其中：

$$Z(x) = \alpha_n^{T}(x) \cdot \mathbf{1}$$

条件随机场的概率计算问题

- 期望值的计算
- 利用前向-后向向量，可以计算特征函数关于联合分布P(X,Y)和条件分布P(Y|X)的数学期望。
- 特征函数f_k关于条件分布P(Y|X)的数学期望是：

$$E_{P(Y|x)}[f_k]=\sum_y P(y|x)f_k(y,x)$$

$$=\sum_{i=1}^{n+1}\sum_{y_{i-1}y_i} f_k(y_{i-1},y_i,x,i)\frac{\alpha_{i-1}^{\mathrm{T}}(y_{i-1}|x)M_i(y_{i-1},y_i|x)\beta_i(y_i|x)}{Z(x)}$$

$$k=1,2,\cdots,K$$

- 其中：

$$Z(x)=\alpha_n^{\mathrm{T}}(x)\cdot\mathbf{1}$$

条件随机场的概率计算问题

- 假设经验分布为：$\tilde{P}(X)$ 特征函数fk关于联合分布P(X,Y)的数学期望是：

$$E_{P(X,Y)}[f_k]$$

$$= \sum_{x,y} P(x,y) \sum_{i=1}^{n+1} f_k(y_{i-1}, y_i, x, i)$$

$$= \sum_{x} \tilde{P}(x) \sum_{y} P(y \mid x) \sum_{i=1}^{n+1} f_k(y_{i-1}, y_i, x, i)$$

$$= \sum_{x} \tilde{P}(x) \sum_{i=1}^{n+1} \sum_{y_{i-1} y_i} f_k(y_{i-1}, y_i, x, i) \frac{\alpha_{i-1}^{\mathrm{T}}(y_{i-1} \mid x) M_i(y_{i-1}, y_i \mid x) \beta_i(y_i \mid x)}{Z(x)}$$

$$Z(x) = \alpha_n^{\mathrm{T}}(x) \cdot \mathbf{1}$$

四、条件随机场的学习算法

- 改进的迭代尺度法：
- 已知训练数据集，可知经验分布：$\tilde{P}(X,Y)$ 可通过极大化训练数据的对数似然函数来求模型参数：
- 似然函数：

$$L(w)=L_{\tilde{P}}(P_w)=\log\prod_{x,y}P_w(y|x)^{\tilde{P}(x,y)}=\sum_{x,y}\tilde{P}(x,y)\log P_w(y|x)$$

- 当P为条件随机场模型时：

$$\begin{aligned}L(w)&=\sum_{x,y}\tilde{P}(x,y)\log P_w(y|x)\\&=\sum_{x,y}\left[\tilde{P}(x,y)\sum_{k=1}^{K}w_k f_k(y,x)-\tilde{P}(x,y)\log Z_w(x)\right]\\&=\sum_{j=1}^{N}\sum_{k=1}^{K}w_k f_k(y_j,x_j)-\sum_{j=1}^{N}\log Z_w(x_j)\end{aligned}$$

条件随机场的学习算法

- 改进的迭代尺度法：
- 不断优化对数似然函数改变量的下界：
- 假设模型当前参数向量： $w=(w_1,w_2,\cdots,w_K)^{\mathrm{T}}$
- 向量增量： $\delta=(\delta_1,\delta_2,\cdots,\delta_K)^{\mathrm{T}}$
- 更新向量： $w+\delta=(w_1+\delta_1,w_2+\delta_2,\cdots w_K+\delta_K)^{\mathrm{T}}$
- 关于转移特征t_k的更新方程：

$$
\begin{aligned}
E_{\tilde{P}}[t_k] &= \sum_{x,y}\tilde{P}(x,y)\sum_{i=1}^{n+1}t_k(y_{i-1},y_i,x,i) \\
&= \sum_{x,y}\tilde{P}(x)P(y\mid x)\sum_{i=1}^{n+1}t_k(y_{i-1},y_i,x,i)\exp(\delta_k T(x,y)) \\
& \qquad k=1,2,\cdots,K_1
\end{aligned}
$$

条件随机场的学习算法

- 改进的迭代尺度法：
- 关于转移特征s_l的更新方程：

$$E_{\tilde{p}}[s_l] = \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n+1} s_l(y_i, x, i)$$

$$= \sum_{x,y} \tilde{P}(x) P(y \mid x) \sum_{i=1}^{n} s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x,y))$$

$$l = 1, 2, \cdots, K_2$$

- T(x,y)是在数据(x,y)中出现所有特征数的总和

$$T(x,y) = \sum_k f_k(y,x) = \sum_{k=1}^{K} \sum_{i=1}^{n+1} f_k(y_{i-1}, y_i, x, i)$$

条件随机场的学习算法

- 条件随机场模型学习的改进的迭代尺度法：

输入：特征函数 $t_1, t_2, \cdots, t_{K_1}$， $s_1, s_2, \cdots, s_{K_2}$；经验分布 $\tilde{P}(x, y)$；

输出：参数估计值 $\hat{w}$；模型 $P_{\hat{w}}$.

（1）对所有 $k \in \{1, 2, \cdots, K\}$，取初值 $w_k = 0$

（2）对每一 $k \in \{1, 2, \cdots, K\}$：

（a）当 $k = 1, 2, \cdots, K_1$ 时，令 δ_k 是方程

$$\sum_{x,y} \tilde{P}(x) P(y \mid x) \sum_{i=1}^{n+1} t_k(y_{i-1}, y_i, x, i) \exp(\delta_k T(x, y)) = E_{\tilde{P}}[t_k]$$

的解；

条件随机场的学习算法

- 条件随机场模型学习的改进的迭代尺度法：

当 $k=K_1+l$， $l=1,2,\cdots,K_2$ 时，令 δ_{K_1+l} 是方程

$$\sum_{x,y}\tilde{P}(x)P(y|x)\sum_{i=1}^{n}s_l(y_i,x,i)\exp(\delta_{K_1+l}T(x,y))=E_{\tilde{P}}[s_l]$$

的解，式中 $T(x,y)$ 由式 (11.38) 给出．

（b）更新 w_k 值： $w_k \leftarrow w_k+\delta_k$

（3）如果不是所有 w_k 都收敛，重复步骤(2)．

条件随机场的学习算法

- T(x,y) 表示数据(x,y)中的特征总数，对不同的数据(x,y)取值可能布同，定义松弛特征：

$$s(x,y)=S-\sum_{i=1}^{n+1}\sum_{k=1}^{K}f_k(y_{i-1},y_i,x,i)$$

- S为大的常数，使得对训练数据集所有(x,y)

$$s(x,y)\geqslant 0$$

条件随机场的学习算法

- 对于转移特征：δ_k的更新方程为：

$$\sum_{x,y}\tilde{P}(x)P(y\mid x)\sum_{i=1}^{n+1}t_k(y_{i-1},y_i,x,i)\exp(\delta_k S)=E_{\tilde{P}}[t_k]$$

$$\delta_k=\frac{1}{S}\log\frac{E_{\tilde{P}}[t_k]}{E_P[t_k]}$$

- 其中：

$$E_P(t_k)=\sum_{x}\tilde{P}(x)\sum_{i=1}^{n+1}\sum_{y_{i-1},y_i}t_k(y_{i-1},y_i,x,i)\frac{\alpha_{i-1}^{\mathrm{T}}(y_{i-1}\mid x)M_i(y_{i-1},y_i\mid x)\beta_i(y_i\mid x)}{Z(x)}$$

条件随机场的学习算法

- 对于状态特征：δ_k的更新方程为：

$$\sum_{x,y}\tilde{P}(x)P(y\,|\,x)\sum_{i=1}^{n}s_l(y_i,x,i)\exp(\delta_{K_1+l}S)=E_{\tilde{P}}[s_l]$$

$$\delta_{K_1+l}=\frac{1}{S}\log\frac{E_{\tilde{P}}[s_l]}{E_P[s_l]}$$

- 其中：

$$E_P(s_l)=\sum_{x}\tilde{P}(x)\sum_{i=1}^{n}\sum_{y_i}s_l(y_i,x,i)\frac{\alpha_i^T(y_i\,|\,x)\beta_i(y_i\,|\,x)}{Z(x)}$$

- 因担心S过大，每个观测序列x计算其特征最大值
- $T(x)=\max_y T(x,y)$-后向公式计算T(x)=t

条件随机场的学习算法

- 关于转移特征参数的更新方程可以写成：

$$
\begin{aligned}
E_{\tilde{P}}[t_k] &= \sum_{x,y} \tilde{P}(x) P(y \mid x) \sum_{i=1}^{n+1} t_k(y_{i-1}, y_i, x, i) \exp(\delta_k T(x)) \\
&= \sum_{x} \tilde{P}(x) \sum_{y} P(y \mid x) \sum_{i=1}^{n+1} t_k(y_{i-1}, y_i, x, i) \exp(\delta_k T(x)) \\
&= \sum_{x} \tilde{P}(x) a_{k,t} \exp(\delta_k \cdot t) \\
&= \sum_{t=0}^{T_{\max}} a_{k,t} \beta_k^{t}
\end{aligned}
$$

$a_{k,t}$ 是特征 t_k 的期待值，$\delta_k = \log \beta_k$ 。β_k 是多项式方程唯一的实根

条件随机场的学习算法

- 关于状态特征的参数更新方程可以写成：

$$E_{\tilde{P}}[s_l] = \sum_{x,y} \tilde{P}(x) P(y|x) \sum_{i=1}^{n} s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x))$$

$$= \sum_{x} \tilde{P}(x) \sum_{y} P(y|x) \sum_{i=1}^{n} s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x))$$

$$= \sum_{x} \tilde{P}(x) b_{l,t} \exp(\delta_k \cdot t)$$

$$= \sum_{t=0}^{T_{\max}} b_{l,t} \gamma_l^t$$

$b_{l,t}$ 是特征 s_l 的期望值， $\delta_l = \log \gamma_l$ ， γ_l 是多项式方程唯一的实根

条件随机场的学习算法

- 拟牛顿法：

$$P_w(y \mid x) = \frac{\exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)}{\sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right)}$$

- 学习的优化目标函数：

$$\min_{w \in \mathbf{R}^n} f(w) = \sum_x \tilde{P}(x) \log \sum_y \exp\left(\sum_{i=1}^{n} w_i f_i(x,y)\right) - \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i(x,y)$$

- 梯度函数：

$$g(w) = \sum_{x,y} \tilde{P}(x) P_w(y \mid x) f(x,y) - E_{\tilde{P}}(f)$$

条件随机场的学习算法

- 条件随机场模型学习的BFGS算法

输入：特征函数 $f_1, f_2, \cdots, f_n$；经验分布 $\tilde{P}(X,Y)$；

输出：最优参数值 $\hat{w}$；最优模型 $P_{\hat{w}}(y \mid x)$．

（1）选定初始点 $w^{(0)}$，取 $\boldsymbol{B}_0$ 为正定对称矩阵，置 $k=0$

（2）计算 $g_k = g(w^{(k)})$．若 $g_k = 0$，则停止计算；否则转 (3)

（3）由 $B_k p_k = -g_k$ 求出 p_k

（4）一维搜索：求 λ_k 使得

$$f(w^{(k)} + \lambda_k p_k) = \min_{\lambda \geq 0} f(w^{(k)} + \lambda p_k)$$

条件随机场的学习算法

- 条件随机场模型学习的BFGS算法

（5）置 $w^{(k+1)} = w^{(k)} + \lambda_k p_k$

（6）计算 $g_{k+1} = g(w^{(k+1)})$，

若 $g_k = 0$，则停止计算；否则，按下式求出 B_{k+1}：

$$B_{k+1} = B_k + \frac{y_k y_k^{\mathrm{T}}}{y_k^{\mathrm{T}} \delta_k} - \frac{B_k \delta_k \delta_k^{\mathrm{T}} B_k}{\delta_k^{\mathrm{T}} B_k \delta_k}$$

$$y_k = g_{k+1} - g_k, \quad \delta_k = w^{(k+1)} - w^{(k)}$$

（7）置 $k = k+1$，转 (3)

五、条件随机场的预测算法

- 预测算法：
- 给定条件随机场P(Y|X)和输入序列(观测序列)x，
- 求：条件概率最大的输出序列(标记序列)y*，
- 维特比算法：
- 由：

$$P_w(y|x)=\frac{\exp\left(w\cdot F(y,x)\right)}{Z_w(x)}$$

$$\begin{aligned}y^* &= \arg\max_y P_w(y|x)\\ &= \arg\max_y \frac{\exp(w\cdot F(y,x))}{Z_w(x)}\\ &= \arg\max_y \exp(w\cdot F(y,x))\\ &= \arg\max_y (w\cdot F(y,x))\end{aligned}$$

非规范化概率最大的最优路径

$$\max_y \ (w\cdot F(y,x))$$

条件随机场的预测算法

- 路径表示标记序列：

$$w=(w_1,w_2,\cdots,w_K)^{\mathrm{T}}$$

$$F(y,x)=(f_1(y,x),f_2(y,x),\cdots,f_K(y,x))^{\mathrm{T}}$$

$$f_k(y,x)=\sum_{i=1}^{n} f_k(y_{i-1},y_i,x,i),\quad k=1,2,\cdots,K$$

- 只计算非规范化概率：

$$\max_{y}\quad \sum_{i=1}^{n} w\cdot F_i(y_{i-1},y_i,x)$$

$$F_i(y_{i-1},y_i,x)=(f_1(y_{i-1},y_i,x,i),f_2(y_{i-1},y_i,x,i),\cdots,f_K(y_{i-1},y_i,x,i))^{\mathrm{T}}$$

- 为局部特征向量

条件随机场的预测算法

- 维特比算法：
- 首先求出位置1的各个标记j=1,2..m的非规范化概率：

$$\delta_1(j) = w \cdot F_1(y_0 = start, y_1 = j, x), \quad j = 1, 2, \cdots, m$$

- 由递推公式，求出到位置i的各个标记l=1,2...m的非规范化概率的最大值，同时记录最大值路径：

$$\delta_i(l) = \max_{1 \leq j \leq m} \{\delta_{i-1}(j) + w \cdot F_i(y_{i-1} = j, y_i = l, x)\}, \quad l = 1, 2, \cdots, m$$

$$\Psi_i(l) = \arg\max_{1 \leq j \leq m} \{\delta_{i-1}(j) + w \cdot F_i(y_{i-1} = j, y_i = l, x)\}, \quad l = 1, 2, \cdots, m$$

条件随机场的预测算法

- 维特比算法：
- 直到i=n时终止，这时求得非规范化概率的最大值为：

$$\max_{y}(w\cdot F(y,x))=\max_{1\leqslant j\leqslant m}\delta_n(j)$$

- 及最优路径的终点：

$$y_n^*=\arg\max_{1\leqslant j\leqslant m}\delta_n(j)$$

- 由此最优路径终点返回：

$$y_i^*=\Psi_{i+1}(y_{i+1}^*),\quad i=n-1,n-2,\cdots,1$$

- 得最优路径：

$$y^*=(y_1^*,y_2^*,\cdots,y_n^*)^{\mathrm{T}}$$

条件随机场的预测算法

- 条件随机场预测的维特比算法：

输入：模型特征向量 $F(y,x)$ 和权值向量 w

观测序列 $x=(x_1,x_2,\cdots,x_n)$

输出：最优路径 $y^*=(y_1^*,y_2^*,\cdots,y_n^*)$

（1）初始化

$$\delta_1(j)=w\cdot F_1(y_0=start,y_1=j,x)\,,\quad j=1,2,\cdots,m$$

（2）递推．对 $i=2,3,\cdots,n$

$$\delta_i(l)=\max_{1\leqslant j\leqslant m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\}\,,\quad l=1,2,\cdots,m$$

$$\Psi_i(l)=\arg\max_{1\leqslant j\leqslant m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\}\,,\quad l=1,2,\cdots,m$$

条件随机场的预测算法

- 条件随机场预测的维特比算法：

（3）终止

$$\max_{y}(w \cdot F(y,x)) = \max_{1 \leq j \leq m} \delta_n(j)$$

$$y_n^* = \arg\max_{1 \leq j \leq m} \delta_n(j)$$

（4）返回路径

$$y_i^* = \Psi_{i+1}(y_{i+1}^*), \quad i = n-1, n-2, \cdots, 1$$

求得最优路径 $y^* = (y_1^*, y_2^*, \cdots, y_n^*)$

条件随机场的预测算法

- 例：用维特比算法求给定输入序列（观测序列）x对于的最优输出序列（标记序列 $y^* = (y_1^*, y_2^*, y_3^*)$
- 利用维特比法求最优路径问题：

$$\max \sum_{i=1}^{3} w \cdot F_i(y_{i-1}, y_i, x)$$

（1）初始化

$$\delta_1(j) = w \cdot F_1(y_0 = start, y_1 = j, x), \quad j = 1,2$$

$$i = 1, \quad \delta_1(1) = 1, \quad \delta_1(2) = 0.5$$

（2）递推

$$i = 2 \quad \delta_2(l) = \max_j \{\delta_1(j) + w \cdot F_2(j, l, x)\}$$

$$\delta_2(1) = \max\{1 + \lambda_2 t_2, 0.5 + \lambda_4 t_4\} = 1.6, \quad \Psi_2(1) = 1$$

$$\delta_2(2) = \max\{1 + \lambda_1 t_1 + \mu_2 s_2, 0.5 + \mu_2 s_2\} = 2.5, \quad \Psi_2(2) = 1$$

条件随机场的预测算法

$i=3$ $\delta_3(l)=\max_j\{\delta_2(j)+w\bullet F_3(j,l,x)\}$

$\delta_3(1)=\max\{1.6+\mu_5 s_5, 2.5+\lambda_3 t_3+\mu_3 s_3\}=4.3$， $\Psi_3(1)=2$

$\delta_3(2)=\max\{1.6+\lambda_1 t_1+\mu_4 s_4, 2.5+\lambda_5 t_5+\mu_4 s_4\}=3.2$， $\Psi_3(2)=1$

（3）终止

$$\max_y(w\bullet F(y,x))=\max\delta_3(l)=\delta_3(1)=4.3$$

$$y_3^*=\arg\max_l\delta_3(l)=1$$

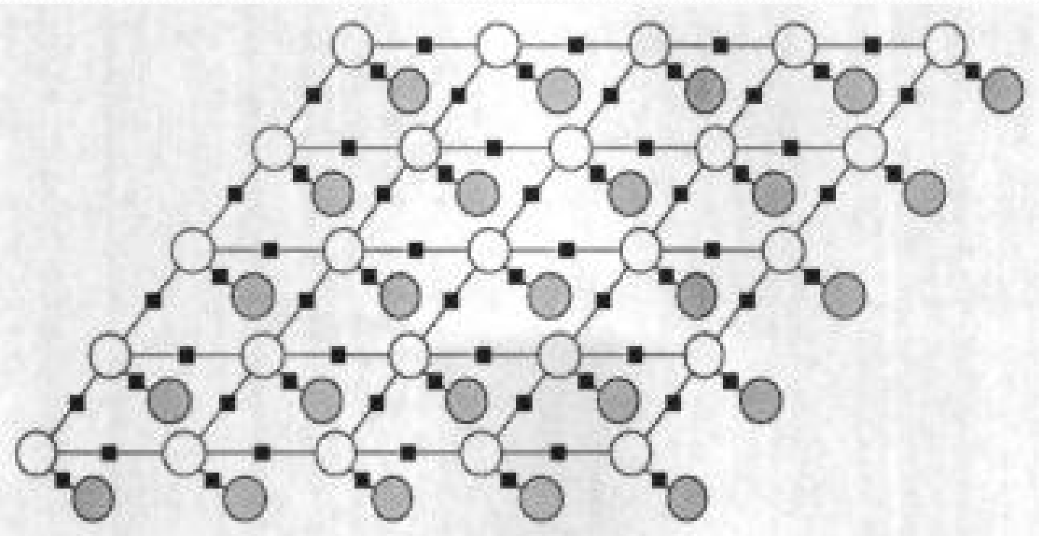
（4）返回

$$y_2^*=\Psi_3(y_3^*)=\Psi_3(1)=2$$

$$y_1^*=\Psi_2(y_2^*)=\Psi_2(2)=1$$

最优标记序列 $y^*=(y_1^*,y_2^*,y_3^*)=(1,2,1)$

基于二维条件随机场的视频分割



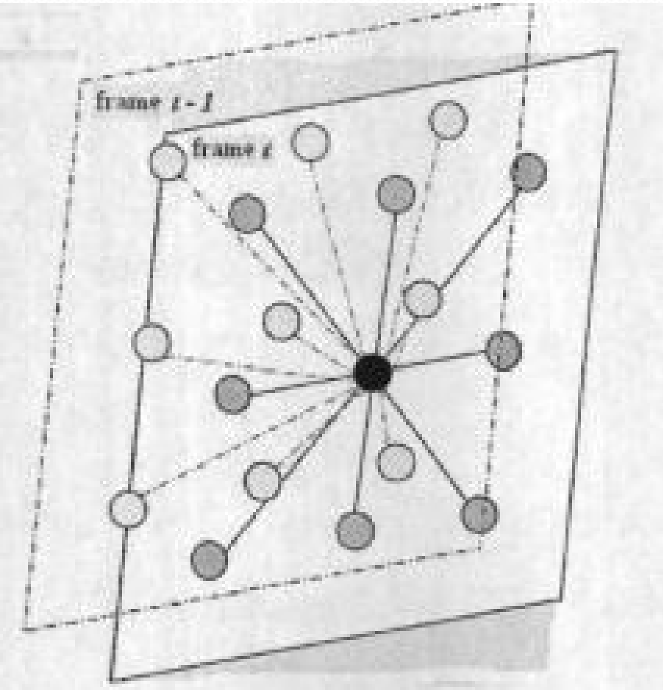


图 3-3 视频序列中的时空邻域关系

基于二维条件随机场的视频分割

- 标签： $L_i = \{0,1,2\}$
- 二维条件随机场模型：

$$P(L \mid X) = \frac{1}{Z(X)} \exp(-E(L;X))$$

$$E(L;X) = \sum_i \lambda_i f_1(L_i;X_i) + \nu \sum_{j \in N(i) \cup M(i)} f_2(L_i, L_j)$$

$$f_1(L_i;X_i) = \delta(L_i, L_{i,m})$$

$$\lambda_i = \begin{cases} \log P(X_i \mid L_i = 0) & if \ L_i = 0 \\ \log P(X_i \mid L_i = 1) & if \ L_i = 1 \\ \log P(X_i \mid L_i = 2) & else \end{cases}$$

基于二维条件随机场的视频分割

$$f_2(L_i, L_j) = \begin{cases} \beta_1 & if\ L_i = L_j = 0 \\ \beta_2 & if\ L_i = L_j = 1 \\ \beta_3 & if\ L_i = L_j = 2 \\ \beta_4 & if\ L_i \neq L_j \end{cases}$$

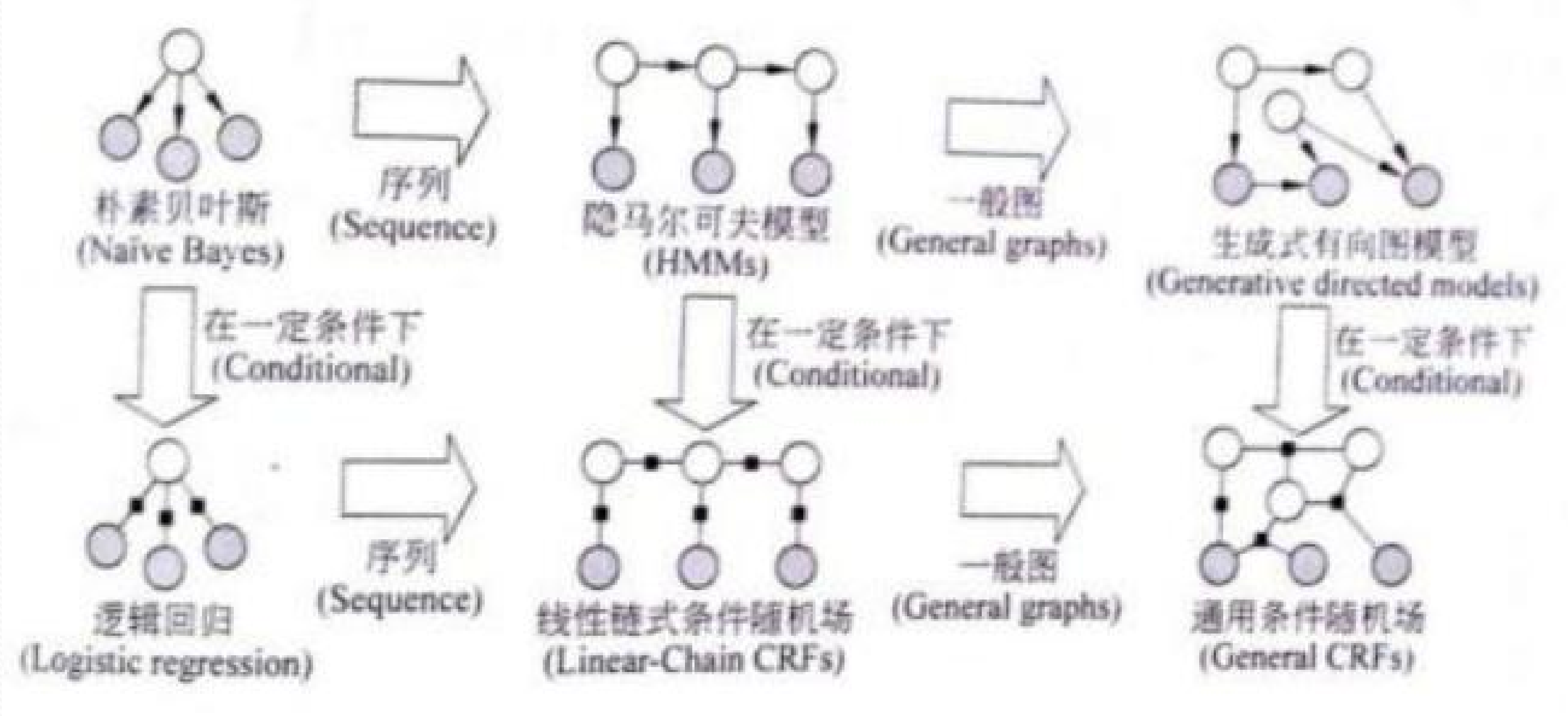
$$f_2(L_i^t, L_j^{t-1}) = \begin{cases} \phi_1 & if\ L_i^t = L_j^{t-1} = 0 \\ \phi_2 & if\ L_i^t = L_j^{t-1} = 1 \\ \phi_3 & if\ L_i^t = L_j^{t-1} = 2 \\ \phi_4 & if\ L_i^t \neq L_j^{t-1} \end{cases}$$

基于二维条件随机场的视频分割

基于二维条件随机场模型的视频阈值化流程：

1. 初始化条件随机场模型参数 ν，以及常量 β_1、β_2、β_3、β_4、ϕ_1、ϕ_2、ϕ_3 和 ϕ_4；
2. 获取一帧图像，利用背景模型、阴影模型以及前景模型获得帧中各像素对应的分类标签，并计算条件随机场模型参数 λ_1；
3. 根据第 2 步的像素分类标签计算条件随机场模型的各个特征函数；
4. 根据模型参数，计算式子（3.20）获得最优的分割结果；
5. 重复第 2 步直到结束。

模型关联



朴素贝叶斯
(Naive Bayes)
序列
(Sequence)
隐马尔可夫模型
(HMMs)
一般图
(General graphs)
生成式有向图模型
(Generative directed models)
在一定条件下
(Conditional)
在一定条件下
(Conditional)
在一定条件下
(Conditional)
逻辑回归
(Logistic regression)
序列
(Sequence)
线性链式条件随机场
(Linear-Chain CRFs)
一般图
(General graphs)
通用条件随机场
(General CRFs)

Q & A