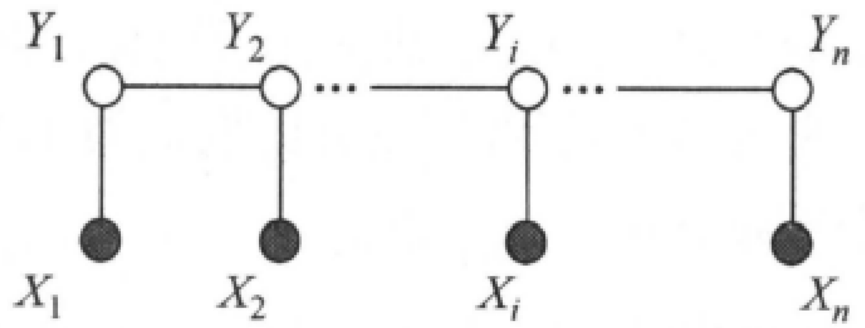
前情回顾

条件随机场CRF

※ 给定一组输入随机变量条件下另一组输出随机变量的条件概率分布模型

- ❖ $P(y|x)$
- ❖ y符合马尔可夫随机场
- ❖ $P\left(y_v \middle| x, y_{V\backslash\{v\}}\right) = P(y_v|x, y_{n\{v\}})$



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

条件随机场CRF

※ $P(y|x)=\frac{1}{Z}\exp\left(\sum_j\sum_1^{n-1}\lambda_j t_j(y_{i+1},y_i,x,i)+\sum_k\sum_{i=1}^{n}\mu_k s_k(y_i,x,i)\right)$

※ $Z=\sum_y\exp\left(\sum_j\sum_1^{n-1}\lambda_j t_j(y_{i+1},y_i,x,i)+\sum_k\sum_{i=1}^{n}\mu_k s_k(y_i,x,i)\right)$

$$\boxed{p(y|x)=\frac{1}{Z}\exp\sum_{k=1}^{K}w_k f_k(y,x)\qquad Z=\sum_y\exp\sum_k w_k f_k}$$

※ 引进起点标记$y_o=start$和终点标记$y_{n+1}=stop$

※ 定义m阶矩阵随机变量：$M_i(x)=[M_i(y_{i-1},y_i|x)]\in R^{m\times m}$(假设y有m个取值)

※ $M_i(y_{i-1},y_i|x)=exp\left(\sum_{k=1}^{K}w_k f_k(y_{i-1},y_i,x,i)\right)$

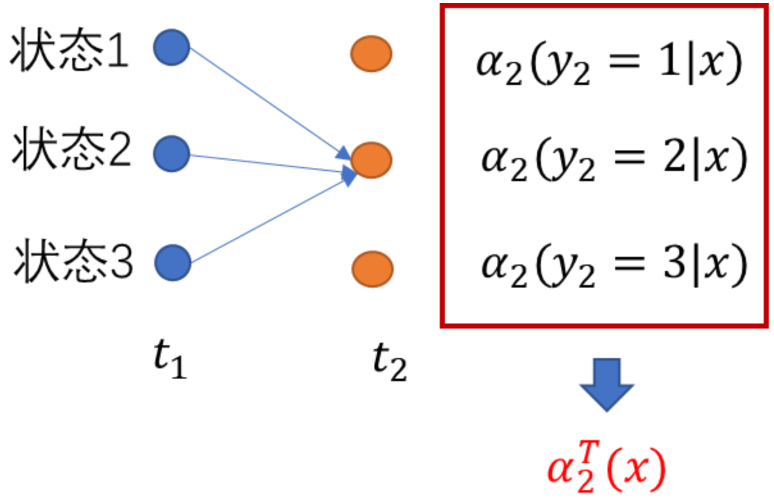
$$\boxed{P(y|x)=\frac{1}{Z}\prod_{i=1}^{n+1}M_i(y_{i-1},y_i|x)\qquad Z=[M_1(x)M_2(x)\cdots M_{n+1}(x)]_{start,stop}}$$

条件随机场CRF

※ 概率计算

❖ 前项向量： $\alpha_i^T(y_i|x) = \alpha_{i-1}^T(y_{i-1}|x)[M_i(y_{i-1}, y_i|x)]$

❖ 后向向量： $\beta_i(y_i|x) = [M_{i+1}(y_i, y_{i+1}|x)]\beta_{i+1}(y_{i+1}|x)$



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

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

※ 参数学习

※ 梯度下降法/拟牛顿法：

※ $\log P(y|x) = \sum_{j=1}^{n}\left[\sum_{k=1}^{K} w_k f_k\left(y_j, x_j\right) - \log Z\left(x_j\right)\right]$

※ $loss = -\log P(y \mid x)$

※ 改进的迭代尺度法

※ $\boldsymbol{max\ L(w)} = \sum_{x,y} \widetilde{P}(x,y) \sum_{i=1}^{n} w_i f_i(x,y) - \sum_x \widetilde{P}(x) \log Z_w(x)$

※ 基本想法：基于当前参数向量，寻找新向量，使得模型的对数似然函数值增大$L(w+\delta) - L(w) \geq A(\delta|w) \geq B(\delta|w)$

※ 预测算法

※ 维特比算法：求非规范化概率最大的最优路径问题

初始

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

递推

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

$$\psi_i(l) = \arg\max_{1\le j\le m}\{\delta_{i-1}(j) + w \cdot F_i(y_{i-1} = j, y_i = l, x)\}$$

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

		start	n	v	p
Bob	n	0.9	—	—	—
	v	0	—	—	—
	p	0	—	—	—

	函数条件	权重
t1	=1 （$y_{t-1}=n$，$y_t=v$） =0 其它	0.6
t2	=1 （$y_{t-1}=p$，$y_t=n$） =0 其它	0.8
t3	=1 （$y_{t-1}=v$，$y_t=n$） =0 其它	0.5
s1	=1 （ $y_t=n$，$x_t=$ 人名 ） =0 其它	0.9
s2	=1 （ $y_t=n$，$x_t=$ 地点 ） =0 其它	0.9
s3	=1 （$y_t=p$，$x_t=at$） =0 其它	0.7

Bob drank coffee at Starbucks

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

		start	n	v	p
Bob	n	0.9	—	—	—
	v	0	—	—	—
	p	0	—	—	—
drank	n	—	0.9	0+0.5=0.5	0+0.8=0.8
	v	—	0.9+0.6=1.5	0	0
	p	—	0.9	0	0

	函数条件	权重
t1	=1 （$y_{t-1}=n$，$y_t=v$） =0 其它	0.6
t2	=1 （$y_{t-1}=p$，$y_t=n$） =0 其它	0.8
t3	=1 （$y_{t-1}=v$，$y_t=n$） =0 其它	0.5
s1	=1 （ $y_t=n$，$x_t=$ 人名 ） =0 其它	0.9
s2	=1 （ $y_t=n$，$x_t=$ 地点 ） =0 其它	0.9
s3	=1 （$y_t=p$，$x_t=at$） =0 其它	0.7

Bob drank coffee at Starbucks

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

		start	n	v	p
Bob	n	0.9	—	—	—
	v	0	—	—	—
	p	0	—	—	—
drank	n	—	0.9	0+0.5=0.5	0+0.8=0.8
	v	—	0.9+0.6=1.5	0	0
	p	—	0.9	0	0
coffee	n	—	0.9	1.5+0.5=2	0.9+0.8=1.7
	v	—	0.9+0.6=1.5	1.5	0.9
	p	—	0.9	1.5	0.9

	函数条件	权重
t1	=1 $(y_{t-1}=n, \ y_t=v)$ =0 其它	0.6
t2	=1 $(y_{t-1}=p, \ y_t=n)$ =0 其它	0.8
t3	=1 $(y_{t-1}=v, \ y_t=n)$ =0 其它	0.5
s1	=1 ($y_t=n, \ x_t=$ 人名) =0 其它	0.9
s2	=1 ($y_t=n, \ x_t=$ 地点) =0 其它	0.9
s3	=1 $(y_t=p, \ x_t=at)$ =0 其它	0.7

Bob drank coffee at Starbucks

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

		start	n	v	p
Bob	n	0.9	—	—	—
	v	0	—	—	—
	p	0	—	—	—
drank	n	—	0.9	0+0.5=0.5	0+0.8=0.8
	v	—	0.9+0.6=1.5	0	0
	p	—	0.9	0	0
coffee	n	—	0.9	1.5+0.5=2	0.9+0.8=1.7
	v	—	0.9+0.6=1.5	1.5	0.9
	p	—	0.9	1.5	0.9
at	n	—	2	1.5+0.5=2	1.5+0.8=2.3
	v	—	2+0.6=2.6	1.5	1.5
	p	—	2+0.7=2.7	1.5+0.7=2.2	1.5+0.7=2.2

	函数条件	权重
t1	=1 $(y_{t-1}=n,\ y_t=v)$ =0 其它	0.6
t2	=1 $(y_{t-1}=p,\ y_t=n)$ =0 其它	0.8
t3	=1 $(y_{t-1}=v,\ y_t=n)$ =0 其它	0.5
s1	=1 （ $y_t=n,\ x_t=$ 人名 ） =0 其它	0.9
s2	=1 （ $y_t=n,\ x_t=$ 地点 ） =0 其它	0.9
s3	=1 $(y_t=p,\ x_t=at)$ =0 其它	0.7

Bob drank coffee at Starbucks

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

		start	n	v	p
Bob **n**	n	0.9	—	—	—
	v	0	—	—	—
	p	0	—	—	—
drank **v**	n	—	0.9	0+0.5=0.5	0+0.8=0.8
	v	—	0.9+0.6=1.5	0	0
	p	—	0.9	0	0
coffee **n**	n	—	0.9	1.5+0.5=2	0.9+0.8=1.7
	v	—	0.9+0.6=1.5	1.5	0.9
	p	—	0.9	1.5	0.9
at **p**	n	—	2	1.5+0.5=2	1.5+0.8=2.3
	v	—	2+0.6=2.6	1.5	1.5
	p	—	2+0.7=2.7	1.5+0.7=2.2	1.5+0.7=2.2
starbucks **n**	n	—	2.3+0.9=3.2	2.6+0.5+0.9 =4	2.7+0.8+0.9 =4.4
	v	—	2.3+0.6=2.9	2.6	2.7
	p	—	2.3	2.6	2.7

	函数条件	权重
t1	=1 $(y_{t-1}=n,\ y_t=v)$ =0 其它	0.6
t2	=1 $(y_{t-1}=p,\ y_t=n)$ =0 其它	0.8
t3	=1 $(y_{t-1}=v,\ y_t=n)$ =0 其它	0.5
s1	=1 （ $y_t=n,\ x_t=$ 人名 ） =0 其它	0.9
s2	=1 （ $y_t=n,\ x_t=$ 地点 ） =0 其它	0.9
s3	=1 $(y_t=p,\ x_t=at)$ =0 其它	0.7

Bob drank coffee at Starbucks

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

Thanks.