机器学习

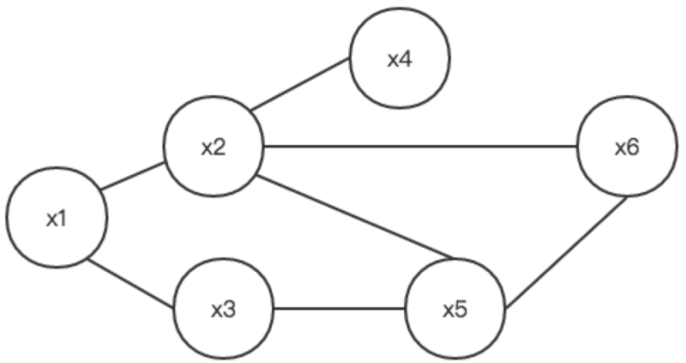
苏州大学计算机科学与技术学院
自然语言处理实验室
主讲：周夏冰
邮箱：zhouxiabing@suda.edu.cn

条件随机场模型

- **条件随机场**(conditional random field, CRF)是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，其特点是假设**输出**随机变量构成马尔可夫随机场 $P(y|x)$
- 马尔可夫随机场：无向图模型

马尔可夫随机场

- 无向图概率图模型：利用无向图结构来表示变量间概率依赖关系
- 无向图$G=(V,E)$
  - **结构**：模型由一组节点（代表随机变量）和连接节点的无向边组成。边表示节点之间存在某种依赖关系，但方向性未定义。
    - 任何结点$\boldsymbol{v} \in \boldsymbol{V}$，表示一个随机变量$\boldsymbol{Y_v}$
    - 边$e \in E$表示随机变量之间的依赖关系



  - 联合概率密度$P(Y)$
    - 通过定义无向图上的势函数（Potential Functions）来描述变量间的联合概率分布

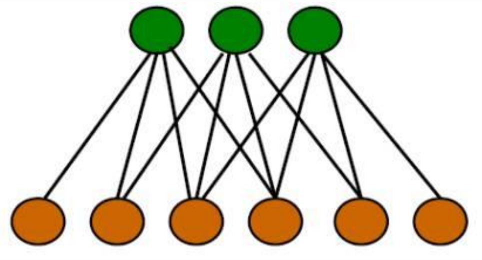
马尔可夫随机场

- 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)
  - 描述随机变量间条件独立性关系
  - 从不同层面刻画了变量间的依赖结构

马尔可夫随机场

成对马尔可夫性

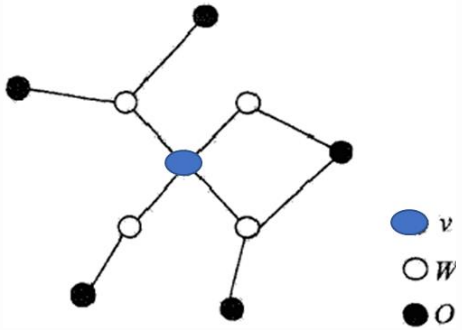
- 设u和v是无向图G中任意两个没有边连接的结点，结点u和v分别对应随机变量Y_u和Y_v，其他所有结点为O，对应的随机变量组是Y_O
- $P(Y_u, Y_v | Y_O) = P(Y_u | Y_O) P(Y_v | Y_O)$



关注非相邻节点的独立性，是局部性质的体现

局部马尔可夫性

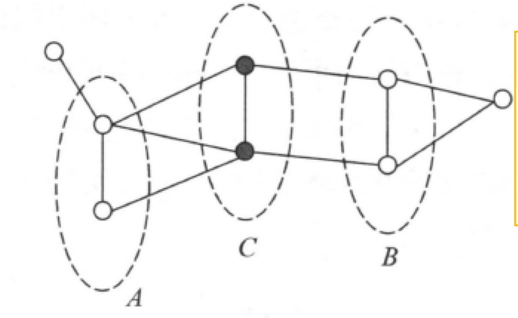
- W是与v有边相连的所有结点，O为其它结点
- $P(Y_v, Y_O | Y_W) = P(Y_v | Y_W) P(Y_O | Y_W)$



强调节点与邻居的依赖关系，是模型推断的基础

全局马尔可夫性

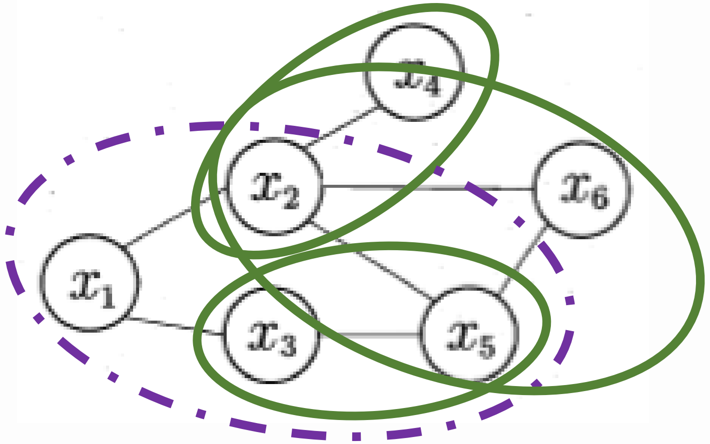
- A和B在无向图中被结点C分隔开
- $P(Y_A, Y_B | Y_C) = P(Y_A | Y_C) P(Y_B | Y_C)$



提供最一般的条件独立性描述，是图模型理论的核心

马尔可夫随机场

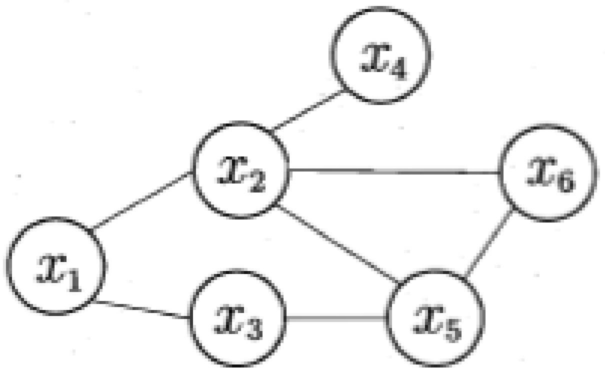
- 无向图的因子分解
  - 团



  - 最大团
    - $\{x_1, x_2\}, \{x_1, x_3\}, \{x_2, x_4\}, \{x_3, x_5\}, \{x_2, x_5, x_6\}$

马尔可夫随机场

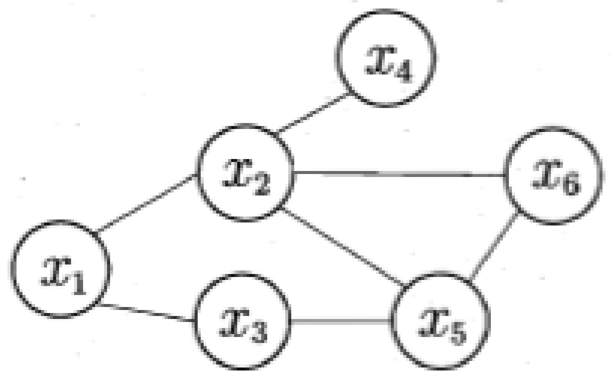
- 无向图的因子分解
  - 团&最大团



  - 势函数
    - 相关关系度量：比如x2与x4相连，则它们具有相关关系$\varphi(x_2, x_4) = \begin{cases} 1.5 & if\ x_2 = x_4 \\ 0.1 & if\ otherwise \end{cases}$

马尔可夫随机场

- 联合概率
  - Hammersley-Clifford定理保证
    - $P(Y) = \frac{1}{Z}\prod_c \psi_c(Y_c)$
    - $Z = \sum_Y \prod_c \psi_c(Y_c)$
  - 基于团分解为多个因子成积
    - $P(x) = \frac{1}{Z}\prod_{Q\in C} \psi_Q(x_Q)$
    - $P(x) = \frac{1}{Z}\psi_{12}(x_1, x_2)\psi_{13}(x_1, x_3)\ \psi_{24}(x_2, x_4)\psi_{35}(x_3, x_5)\ \psi_{256}(x_2, x_5, x_6)$



条件随机场模型

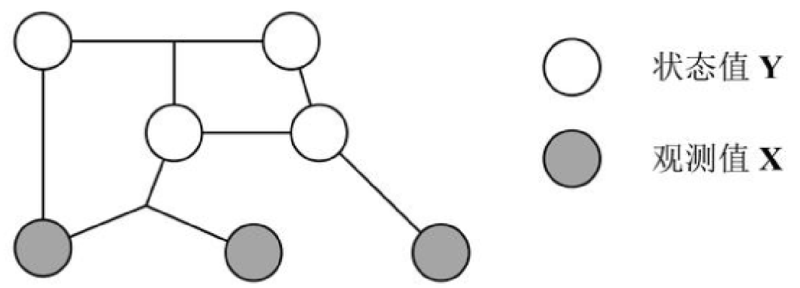
- 条件随机场试图对多个变量在给定观测之后的条件概率进行建模
- $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n\}$ 观测序列
- $\boldsymbol{y} = \{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_n\}$ 为对应的标记序列
- 目标：构建条件概率$\boldsymbol{P}(\boldsymbol{y}|\boldsymbol{x})$

马尔可夫性确保 **y** 的依赖关系仅通过图结构表达，避免全连接导致的计算复杂度爆炸

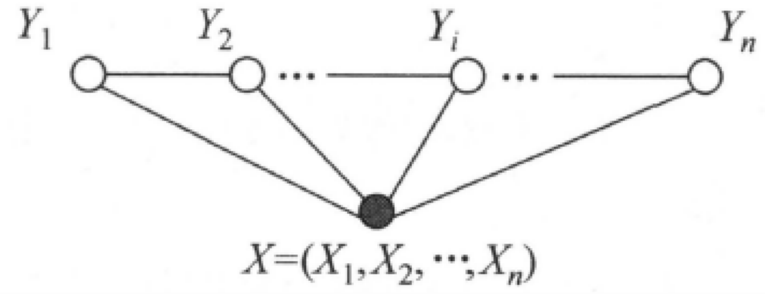
- 图$G = <V, E>$表示结点与标记变量y中元素一一对应的无向图，$\boldsymbol{y}_v$表示与结点v对应的标记变量，$\boldsymbol{n}(\boldsymbol{v})$表示$v$的邻接结点，如果有：

$\boldsymbol{P}\left(\boldsymbol{y}_v \middle| \boldsymbol{x}, \boldsymbol{y}_{V\setminus\{v\}}\right) = \boldsymbol{P}(\boldsymbol{y}_v|\boldsymbol{x}, \boldsymbol{y}_{n\{v\}})$，则$x, y$就构成了一个条件随机场

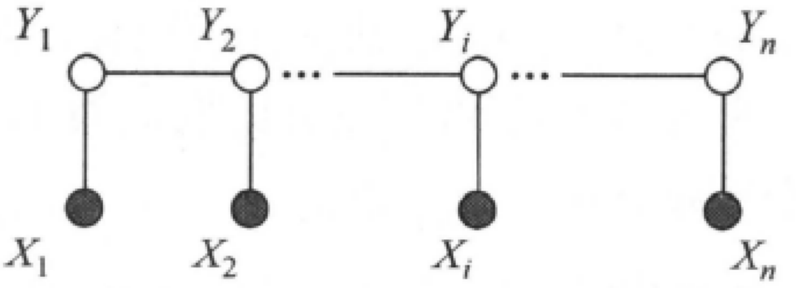
条件随机场模型



- **条件随机场：**
  - 状态y只和相邻状态有关，X不具有马尔科夫性质，作为一个整体影响Y
- **线性条件随机场：**
  - $\boldsymbol{X}=\{\boldsymbol{X_1},\boldsymbol{X_2},\cdots,\boldsymbol{X_n}\}$， $\boldsymbol{Y}=\{\boldsymbol{Y_1},\boldsymbol{Y_2},\cdots,\boldsymbol{Y_n}\}$ 线性链表示的随机变量序列
  - $P(Y_i|X,Y_1,\cdots,Y_{i-1},Y_{i+1},\cdots,Y_n)=P(Y_i|X,Y_{i-1},Y_{i+1})$



(a)



(b)

条件随机场模型

- 参数化形式

转移特征函数　　　　状态特征函数

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

y	$\{y_1$	y_2	y_3	y_4	y_5	$y_6\}$
	D	N	V	P	D	N
x	$\{x_1$	x_2	x_3	x_4	x_5	$x_6\}$

$$t_j(y_{i+1}, y_i, x, i) = \begin{cases} 1, if\ y_{i+1} = P, y_i = V\ and\ x_i = "knock" \\ 0, otherwise \end{cases}$$

$$s_k(y_i, x, i) = \begin{cases} 1, if\ y_i = V\ and\ x_i = "knock" \\ 0, otherwise \end{cases}$$

条件随机场模型

- $P(y|x)=\frac{1}{Z}\exp\left(\boxed{\sum_j}\sum_1^{n-1}\lambda_j t_j(y_{i-1},y_i,x,i)+\boxed{\sum_k}\sum_{i=1}^{n}\mu_k s_k(y_i,x,i)\right)$
- $f_k(y_{i-1},y_i,x,i)=\begin{cases} t_k(y_{i-1},y_i,x,i), & k=1,2,\cdots,K_1 \\ s_k(y_i,x,i), & k=K_1+l, l=1,2\cdots,K_2 \end{cases}$
- $w_k=\begin{cases} \lambda_k, k=1,2,\cdots,K_1 \\ \mu_k, k=K_1+l, l=1,2\cdots,K_2 \end{cases}$

$w=\begin{bmatrix} w^1 \\ w^2 \\ \cdots \\ w^K \end{bmatrix}$ $F=[f_1,f_2,\cdots,f_K]$

简化形式

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

$Z=\sum_y \exp\sum_k w_k f_k$

向量内积形式

$p(y|x)=\frac{\exp(w\cdot F(y,x))}{Z}$

$Z=\sum_y \exp w\cdot F(y,x)$

条件随机场模型

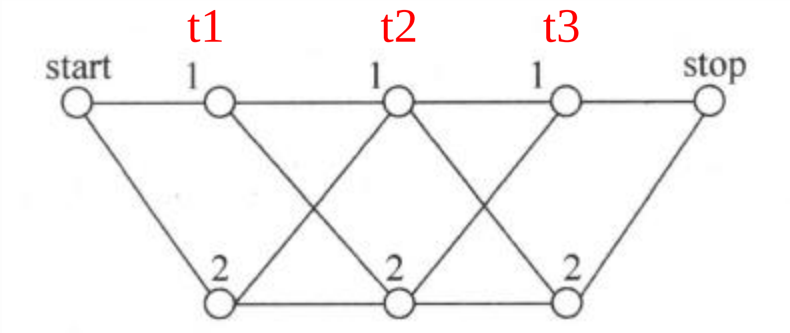
- **矩阵形式**

序列长度

  - 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个状态

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



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

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

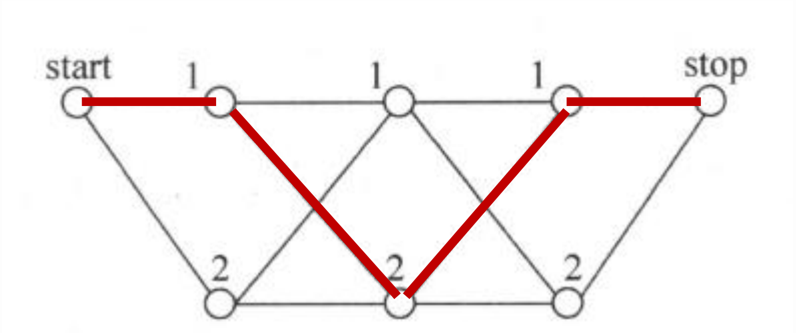
条件随机场模型

- **矩阵形式**
  - 条件概率矩阵形式：

$$\boxed{M_1(start,1|x)M_2(1,2|x)M_3(2,1|x)M_4(1,stop|x)}$$

$\boldsymbol{a_{01}}$ $\boldsymbol{b_{12}}$ $\boldsymbol{c_{21}}$ $\boldsymbol{1}$

    - $P(y|x) = \frac{1}{Z}\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i|x)$
    - $Z = [M_1(x)M_2(x)\cdots M_{n+1}(x)]_{start,stop}$ 从star到stop的所有路径



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

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

条件随机场模型

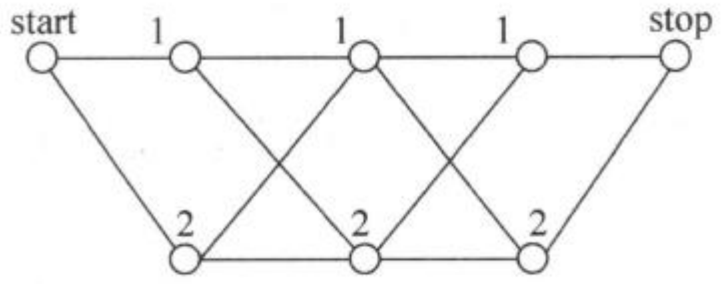
- **矩阵形式**

  - $Z = [M_1(x)M_2(x)\cdots M_{n+1}(x)]_{start,stop}$ 从star到stop的所有路径

$$M_1M_2 = \begin{bmatrix} a_{01}b_{11}+a_{02}b_{21} & a_{01}b_{12}+a_{02}b_{22} \\ 0 & 0 \end{bmatrix}$$

$$M_1M_2M_3 = \begin{bmatrix} a_{01}b_{11}c_{11}+a_{02}b_{21}c_{11}+a_{01}b_{12}c_{21}+a_{02}b_{22}c_{21} & a_{01}b_{11}c_{12}+a_{02}b_{21}c_{12}+a_{01}b_{12}c_{22}+a_{02}b_{22}c_{22} \\ 0 & 0 \end{bmatrix}$$

$$M_1M_2M_3M_4 = \begin{bmatrix} a_{01}b_{11}c_{11}+a_{02}b_{21}c_{11}+a_{01}b_{12}c_{21}+a_{02}b_{22}c_{21}+a_{01}b_{11}c_{12}+a_{02}b_{21}c_{12}+a_{01}b_{12}c_{22}+a_{02}b_{22}c_{22} & 0 \\ 0 & 0 \end{bmatrix}$$



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

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

三个问题

概率计算

- $P(Y_i = y_i|x)$
- $P(Y_i = y_i, Y_{i-1} = y_{i-1}|x)$

学习算法

- 参数估计

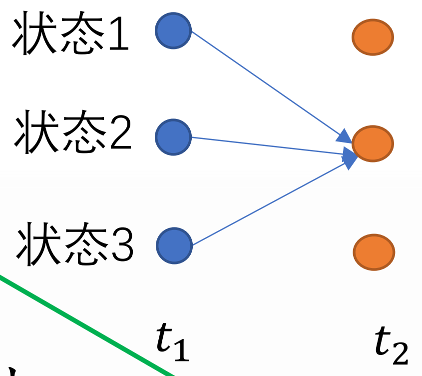
预测算法

- 求最大概率的输出序列

概率计算

$[\alpha_1(y_2=1),\alpha_1(y_2=2),\boxed{\alpha_1(y_2=3)}]$

$$M_2=\begin{bmatrix} M_2(1,1) & M_2(1,2) & M_2(1,3)\\ M_2(2,1) & M_2(2,2) & M_2(2,3)\\ M_2(3,1) & \boxed{M_2(3,2)} & M_2(3,3)\end{bmatrix}$$



$\alpha_2(y_2=1|x)$
$\alpha_2(y_2=2|x)$
$\alpha_2(y_2=3|x)$

$\alpha_2^T(x)$

- **计算**
  - 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：$\boldsymbol{P(Y_i=y_i|x)}$，$\boldsymbol{P(Y_i=y_i,Y_{i-1}=y_{i-1}|x)}$

$$\alpha_2^T(y_2=2|x)=\sum_j \alpha_1^T(y_1=j|x)M_i(j,2|x)$$

- **前向向量$\alpha_i(x)$:**
  - 初始：$\alpha_0(y|x)=\begin{cases}1, & y=start\\ 0, & otherwise\end{cases}$
  - 递推：$\alpha_i^T(x)=\alpha_{i-1}^T(x)M_i(x)$

m维向量（m个状态）

m*m的矩阵

概率计算

- **后向向量$\boldsymbol{\beta_i(x)}$**:

  - 初始：$\beta_{n+1}(y_{n+1}|x) = \begin{cases} 1, & y_{n+1} = stop \\ 0, & otherwise \end{cases}$
  - 递推：$\beta_i(x) = M_{i+1}(x)\beta_{i+1}(x)$
  - $\beta_i(y_i = k|x) = \sum_j [M_{i+1}(k, j|x)]\beta_{i+1}(y_{i+1} = j|x), i = 1,2,\cdots,n$

概率计算

- 位置是i标记为$\boldsymbol{y_i}$的条件概率
  - $P(Y_i = y_i|x) = \dfrac{\alpha_i^T(y_i|x)\beta_i(y_i|x)}{Z(x)}$
- 位置$i-1$标记为$\boldsymbol{y_{i-1}}$，位置i标记是$\boldsymbol{y_i}$的条件概率
  - $P(Y_{i-1} = y_{i-1}, Y_i = y_i|x) = \dfrac{\alpha_{i-1}^T(y_{i-1}|x)M_i(y_{i-1}, y_i|x)\beta_i(y_i|x)}{Z(x)}$
  - $Z(x) = \alpha_n^T(x)\mathbf{1} = \mathbf{1}\beta_1(x)$

概率计算

- 特征函数f_k关于条件分布$P(Y|X)$的数学期望是：
  - $E_{P(Y|X)}[f_k] = \sum_y P(y|x) f_k(y,x)$

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

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

期望

- 设随机变量 X 可以取值为 $x_1, x_2, x_3, \ldots$，对应的概率分别为 $p_1, p_2, p_3, \ldots$，那么，X 的数学期望 $E(X)$ 定义为：
  - $E(X) = \sum_i x_i p_i$
  - 描述随机变量在大量重复实验中可能取得的平均值
  - 对随机变量长期表现的一种度量，反映了随机事件在大量重复发生时的“平均结果”

概率计算

- 特征函数f_k关于条件分布$P(Y|X)$的数学期望是：
  - $E_{P(Y|X)}[f_k] = \sum_y P(y|x) f_k(y,x)$

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

  - 衡量该特征在模型预测分布中“平均活跃程度”
  - 反映了在给定 $P(Y \mid X)$ 下，该特征函数在所有可能的 $\boldsymbol{y}$ 上的平均“激活程度”

假设 x=[“我”,“爱”,“自然语言处理”]，y 是词性标签序列。
$f_k(y_{i-1},y_i,x,i) = I(y_{i-1}=n, y_i=v, x_i=\text{爱})$

1.通过前向-后向算法计算 $P(y_{i-1}=n, y_i=v \mid \boldsymbol{x})$。
2.期望为 $\sum_i P(y_{i-1}=n, y_i=v \mid \boldsymbol{x}) \times 1$（若 $\boldsymbol{x}_i$ =“爱”）。

若期望较高，说明模型认为“名词→动词”的转移在“爱”附近很常见

概率计算

- 特征函数f_k关于联合分布$P(X,Y)$的数学期望是：
  - $E_{P(X,Y)}[f_k] = \sum_{x,y} P(x,y) f_k$
  - 假设 X 是图像，Y 是类别标签。特征函数 $f_k = I[\boldsymbol{y} = \text{“猫”} \wedge \text{图像中有胡须}]$。
  - 联合期望 $E_{P(X,Y)}[f_k]$ 表示“图像中有胡须且类别是猫”的平均概率。
  - 若期望较高，说明模型认为“胡须”是“猫”类别的重要特征。

概率计算

- 特征函数f_k关于联合分布$P(X,Y)$的数学期望是：
  - $P(X,Y)=P(Y|X)\tilde{P}(x)$
  - $E_{P(X,Y)}[f_k]=\sum_{x,y}P(x,y)\sum_{i=1}^{n+1}f_k(y_{i-1},y_i,x,i)$

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

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

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

学习算法

- 条件随机场
  - $p(y|x) = \frac{1}{Z}\exp\sum_{k=1}^{K} w_k f_k(y,x)$
  - $Z = \sum_y exp\sum_{k=1}^{K} w_k f_k(y,x)$
- 学习目标是最大化训练数据上的条件对数似然函数
  - $\log P(y|x) = \sum_{j=1}^{n}\left[\sum_{k=1}^{K} w_k f_k\left(y_j, x_j\right) - \log Z\left(x_j\right)\right]$

学习算法

- 条件随机场
  - $p(y|x) = \frac{1}{Z}\exp\sum_{k=1}^{K} w_k f_k(y,x)$
  - $Z = \sum_y exp\sum_{k=1}^{K} w_k f_k(y,x)$
- 学习目标是最大化训练数据上的条件对数似然函数
  - $\log P(y|x) = \sum_{j=1}^{n}\left[\sum_{k=1}^{K} w_k f_k\left(y_j, x_j\right) - \log Z\left(x_j\right)\right]$
  - 定义训练数据的经验分布为：$\tilde{p}(x,y)$
  - $\boldsymbol{L(w) = \sum_{x,y}\widetilde{P}(x,y)\sum_{i=1}^{n} w_i f_i(x,y) - \sum_{x,y}\widetilde{P}(x,y)\log Z_w(x)}$

学习算法

$$\sum_{x,y} \tilde{P}(x,y) = \sum_{x} \tilde{P}(x)$$

- 条件随机场
  - $\max L(w) = \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i(x,y) - \boxed{\sum_x \tilde{P}(x)} \log Z_w(x)$
  - 梯度下降法：
    - $loss = -\log P(y|x)$
  - 改进的迭代尺度法
    - 显示使用了训练数据的经验分布 $\max L(w) = \max \sum_{x,y} \tilde{P}(x,y) \log P(y|x)$
    - 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

$$\boxed{L(w+\delta) - L(w) \geq A(\delta|w) \geq B(\delta|w)}$$

学习算法

- 条件随机场

	梯度下降	改进的迭代尺度法（IIS）
优化目标	最大化对数条件似然函数	最大化对数条件似然 + 保持特征期望等于经验分布
使用方法	基于梯度信息直接更新参数（一阶导数）	基于约束优化，逐特征更新参数
特征期望	通过梯度差进行调整	显式强制匹配经验分布
是否使用显式的经验分布	隐式体现在梯度中	显式作为迭代中的约束

学习算法

- 条件随机场
  - $\boldsymbol{max\ L(w) = \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n} w_i f_i(x,y) - \sum_x \tilde{P}(x) \log Z_w(x)}$
  - 梯度下降法 $\theta^{t+1} = \theta^t - \eta \nabla L(\theta^t)$
  - 改进的迭代尺度法
  - 拟牛顿法
    - $loss = -\log P(y|x)$
    - 一种高效的二阶优化方法，使用近似的Hessian矩阵信息，提高收敛速度
    - $\theta^{t+1} = \theta^t - \eta H^{-1} \nabla L(\theta^t)$， H^{-1}是梯度变化和参数变化构造出的近似Hessian逆

学习算法

- 条件随机场

- 小规模：**IIS**
- 中等规模、精度优先：**L-BFGS**
- 大数据、在线训练：**SGD / Adam**

	梯度下降	改进的迭代尺度法（IIS）	拟牛顿法
优化目标	最大化对数条件似然函数	最大化对数条件似然 + 保持特征期望等于经验分布	最大化对数条件似然函数
使用方法	基于梯度信息直接更新参数（一阶导数）	基于约束优化，逐特征更新参数	一阶导数 + 拟合二阶导数（近似海森矩阵）
特征期望	通过梯度差进行调整	显式强制匹配经验分布	通过梯度差进行调整
是否使用显式的经验分布	隐式体现在梯度中	显式作为迭代中的约束	隐式使用

预测算法

- 目标： $y^* = \underset{y}{\operatorname{argmax}} P_w(y|x)$

$$= \underset{y}{\operatorname{argmax}} \frac{exp(w \cdot F(y,x))}{Z_w(x)}$$

- **维特比算法**：求**非规范化**概率最大的最优路径问题$\max_{y}(w \cdot F(y,x))$

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

预测算法

- 输入：特征向量$F(y,x)$，权值向量w，观测序列：$x=(x_1,x_2,\cdots,x_n)$
- 输出：最优路径$y^*=(y_1^*,y_2^*,\cdots,y_n^*)$
- 初始
  - $\delta_1(j)=w\cdot F_1(y_0=start,y_1=j,x)$，$j=1,2,\cdots,m$
- 递推
  - $\delta_i(l)=\max\limits_{1\le j\le m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\}$，$l=1,2,\cdots,m$
  - $\psi_i(l)=\arg\max\limits_{1\le j\le m}\{\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\}$，$l=1,2,\cdots,m$

预测算法

- 终止
  - $\max_{y}(w \cdot F(y,x)) = \max_{1 \le j \le m} \delta_n(j)$
  - $y_n^* = \arg\max_{1 \le j \le m} \delta_n(j)$
- 返回路径
  - $y_n^* = \psi_{i+1}(y_{i+1}^*),\ i = n-1, n-2, \cdots, 1$
  - 最优路径：$y^* = (y_1^*, y_2^*, \cdots, y_n^*)$

预测算法

- 例 设有一标注问题，输入观测序列为$X = (X_1, X_2, X_3)$，输出标记序列为$Y = (Y_1, Y_2, Y_3)$， $Y_1, Y_2, Y_3 \in \{1,2\}$
- 特征定义：

$t_1(y_{i-1} = 1, y_i = 2, x, i) = 1 \ \ (\mathrm{i} = 2,3)$	$\lambda_1 = 1$
$t_2(y_1 = 1, y_2 = 1, x, 2) = 1$	$\lambda_2 = 0.6$
$t_3(y_2 = 2, y_3 = 1, x, 3) = 1$	$\lambda_3 = 1$
$t_4(y_1 = 2, y_2 = 1, x, 2) = 1$	$\lambda_4 = 1$
$t_5(y_2 = 2, y_3 = 2, x, 3) = 1$	$\lambda_5 = 0.2$

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

预测算法

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

$t_1(y_{i-1} = 1, y_i = 2, x, i) = 1 \ \ (\mathrm{i} = 2,3)$	$\lambda_1 = 1$
$t_2(y_1 = 1, y_2 = 1, x, 2) = 1$	$\lambda_2 = 0.6$
$t_3(y_2 = 2, y_3 = 1, x, 3) = 1$	$\lambda_3 = 1$
$t_4(y_1 = 2, y_2 = 1, x, 2) = 1$	$\lambda_4 = 1$
$t_5(y_2 = 2, y_3 = 2, x, 3) = 1$	$\lambda_5 = 0.2$

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

  - $\delta_1(1) = w \cdot F_1(y_0 = start, y_1 = 1, x) = 1 \times 1 = 1$
  - $\delta_1(2) = w \cdot F_1(y_0 = start, y_1 = 2, x) = 1 \times 0.5 = 0.5$

预测算法

$t_1(y_{i-1}=1, y_i=2, x, i)=1 \ (i=2,3)$	$\lambda_1=1$
$t_2(y_1=1, y_2=1, x, 2)=1$	$\lambda_2=0.6$
$t_3(y_2=2, y_3=1, x, 3)=1$	$\lambda_3=1$
$t_4(y_1=2, y_2=1, x, 2)=1$	$\lambda_4=1$
$t_5(y_2=2, y_3=2, x, 3)=1$	$\lambda_5=0.2$

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

- 递推

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

➢ $i=2$

$\{1+0.6+0.8, 0.5+1+0.8\}$

$\boxed{\begin{matrix}\delta_1(1)+wF_2(y_1=1, y_2=1, x)\\ \delta_1(2)+wF_2(y_1=2, y_2=1, x)\end{matrix}}$

$\delta_2(1)=\max\limits_{1\le j\le 2}\{\delta_1(j)+w\cdot F_2(y_1=j, y_2=1, x)\}=2.4 \qquad \psi_2(1)=1$

$\delta_2(2)=\max\limits_{1\le j\le 2}\{\delta_1(j)+w\cdot F_2(y_1=j, y_2=2, x)\}=2.5 \qquad \psi_2(2)=1$

$\boxed{\begin{matrix}\delta_1(1)+wF_2(y_1=1, y_2=2, x)\\ \delta_1(2)+wF_2(y_1=2, y_2=2, x)\end{matrix}}$

$\{1+1+0.5, 0.5+0.5\}$

$\boxed{\begin{matrix}\delta_2(1)+wF_3(y_2=1, y_3=1, x)\\ \delta_2(2)+wF_3(y_2=2, y_3=1, x)\end{matrix}}$

$\{2.4+0.8, 2.5+1+0.8\}$

➢ $i=3$

$\delta_3(1)=\max\limits_{1\le j\le 2}\{\delta_2(j)+w\cdot F_3(y_2=j, y_3=1, x)\}=4.3 \qquad \psi_3(1)=2$

$\delta_3(2)=\max\limits_{1\le j\le 2}\{\delta_2(j)+w\cdot F_3(y_2=j, y_3=2, x)\}=3.9$

$\boxed{\begin{matrix}\delta_2(1)+wF_3(y_2=1, y_3=3, x)\\ \delta_2(2)+wF_3(y_2=2, y_3=3, x)\end{matrix}}$

$\{2.4+1+0.5, 2.5+0.2+0.5\} \qquad \psi_3(2)=1$

预测算法

- 终止
  - $\delta_3(1) = \max\limits_{1\le j\le 2}\{\boldsymbol{\delta_2(j)} + \boldsymbol{w}\cdot\boldsymbol{F_3(y_2 = j, y_3 = 1, x)}\} = 4.3$ $\quad\psi_3(1) = 2$
  - $\delta_3(2) = \max\limits_{1\le j\le 2}\{\boldsymbol{\delta_2(j)} + \boldsymbol{w}\cdot\boldsymbol{F_3(y_2 = j, y_3 = 2, x)}\} = 3.9$ $\quad\psi_3(2) = 1$
  - $y_3^* = 1$
- 最优路径
  - $y_2^* = \psi_3(1) = 2$
  - $y_1^* = \psi_2(2) = 1$
  - $y^* = (1,\ 2,\ 1)$

课堂练习

利用条件随机场进行词性标注。假设词性有名词n，动词v，代词p，定义特征函数及相应权重如下：

	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

请对“Bob drank coffee at Starbucks”进行词性标注（给出具体计算过程）