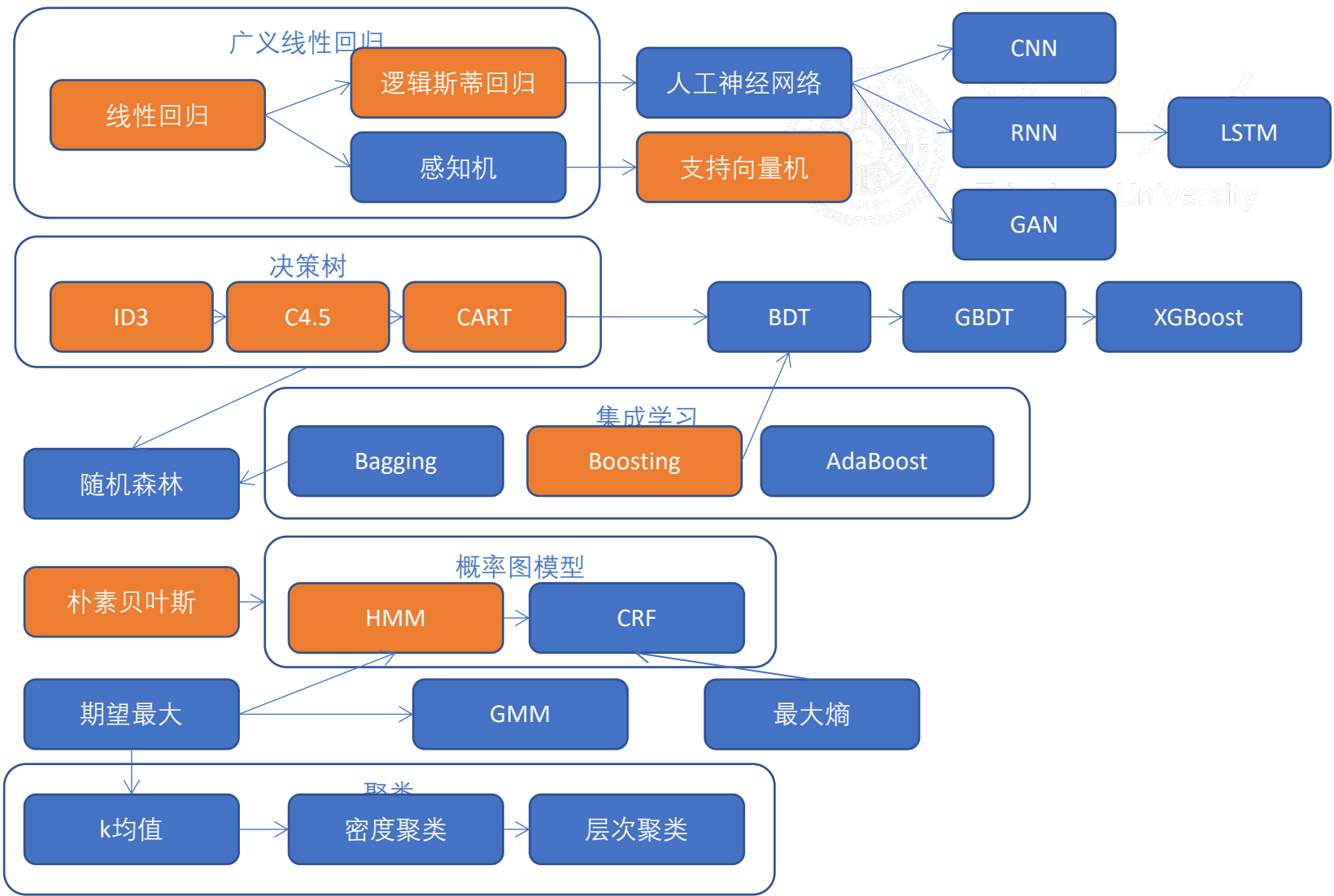


机器学习原理



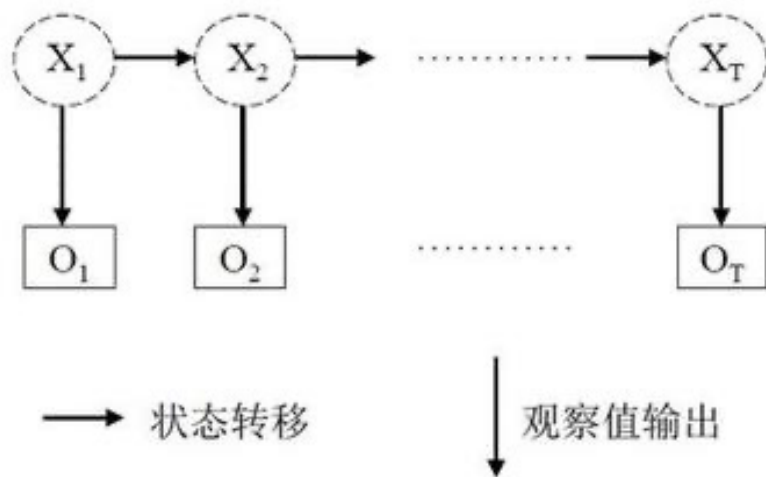


清华大学  
Tsinghua University

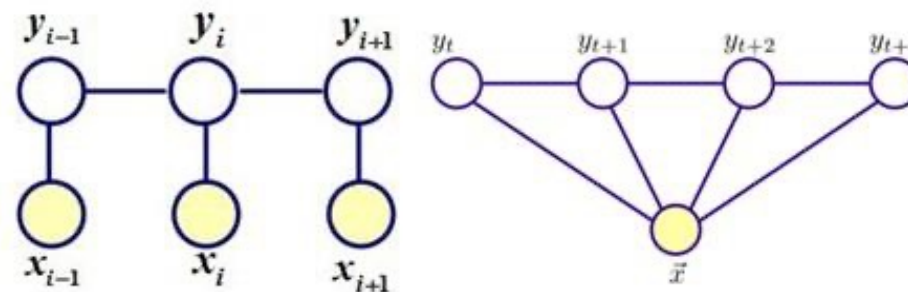
# 第十一章 条件随机场

# HMM到条件随机场

- HMM



- CRF



# HMM和CRF

- 共性：都常用来做序列标注的建模，像词性标注，
- 差异：
  - HMM最大的缺点就是由于其输出独立性假设，导致其不能考虑上下文的特征，限制了特征的选择；在每一节点都要进行归一化，所以只能找到局部的最优值，同时也带来了标记偏见的问题（label bias）；
  - CRF：选择上下文相关特性；不在每一个节点进行归一化，而是所有特征进行全局归一化，可以求得全局的最优值。

# 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。

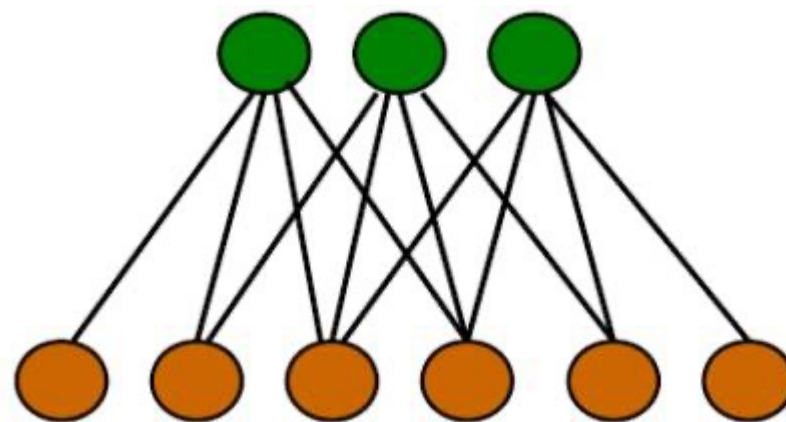
# 模型定义

- 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 $Y_u$ 和 $Y_v$ ，
  - 其他所有结点为 $O$ ，对应的随机变量组是 $Y_O$
  - 给定随机变量组 $Y_O$ 的条件下随机变量 $Y_u$ 和 $Y_v$ 是条件独立的

$$P(Y_u, Y_v | Y_O) = P(Y_u | Y_O)P(Y_v | Y_O)$$





# 模型定义

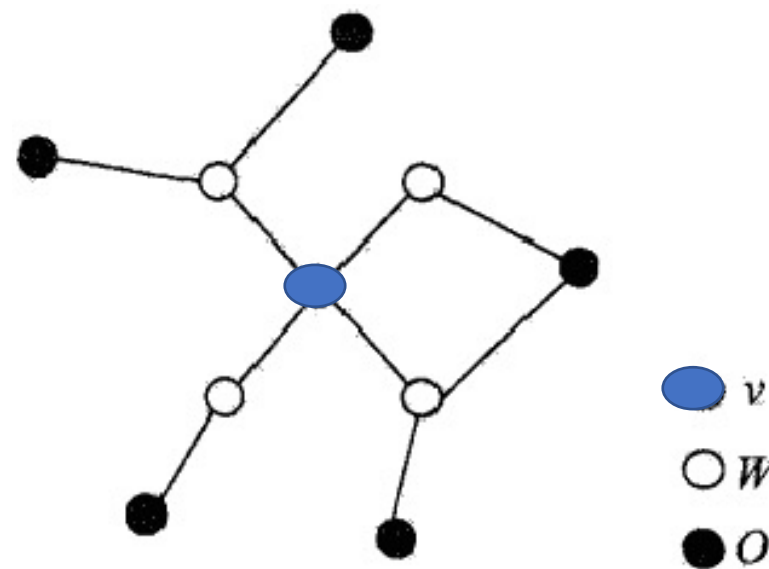
- 局部马尔可夫性( Local Markov properly)

- $v$  任意结点
- $W$ 与 $v$ 有边相连
- $O$  其它

$$P(Y_v, Y_O | Y_W) = P(Y_v | Y_W)P(Y_O | Y_W)$$

- 在  $P(Y_O | Y_W) > 0$  时, 等价于

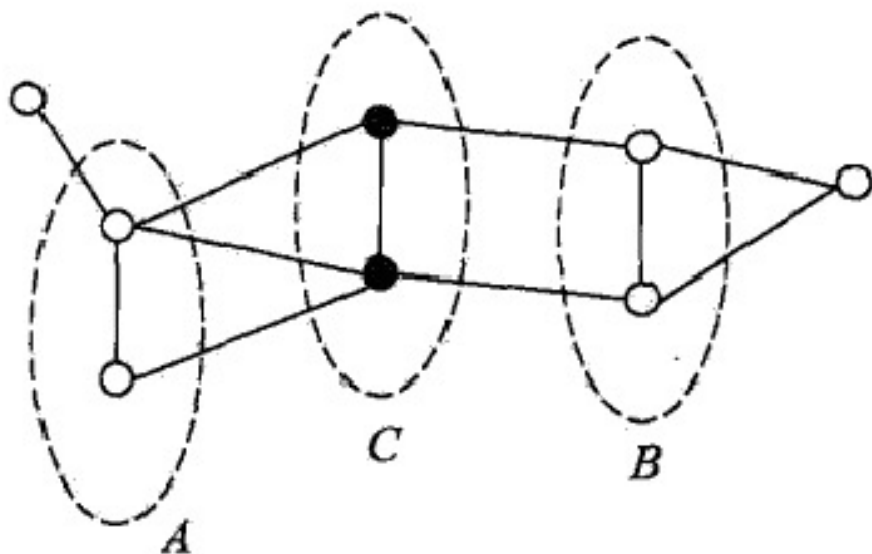
$$P(Y_v | Y_W) = P(Y_v | Y_W, Y_O)$$



# 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,

$$P(Y_A, Y_B | Y_C) = P(Y_A | Y_C)P(Y_B | Y_C)$$



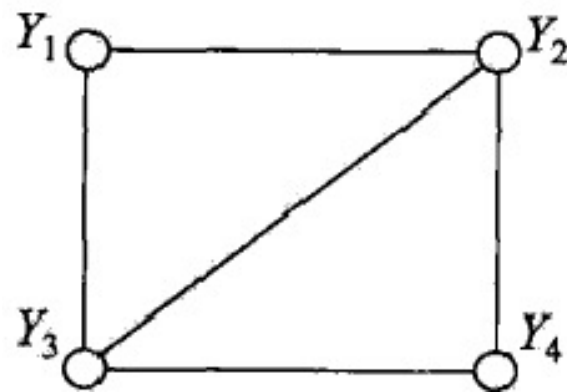
# 模型定义

- 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， $Y_C$ 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：

$$P(Y) = \frac{1}{Z} \prod_c \Psi_c(Y_c)$$

$$Z = \sum_Y \prod_c \Psi_c(Y_c)$$

# 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

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

- 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 $w \sim v$ 表示在图 $G=(V,E)$ 中与结点 $v$ 有边连接的所有结点 $w$ ， $w \neq v$ 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：

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

$$X = (X_1, X_2, \dots, X_n), \quad Y = (Y_1, Y_2, \dots, Y_n)$$

- 最大团是相邻两个结点的集合, 线性链条件随机场:

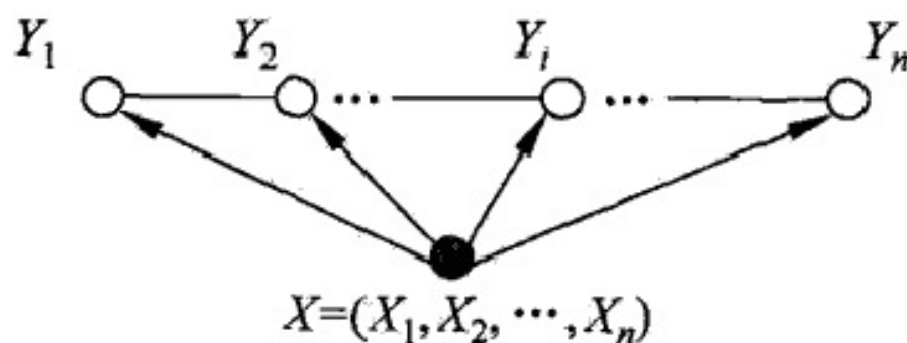
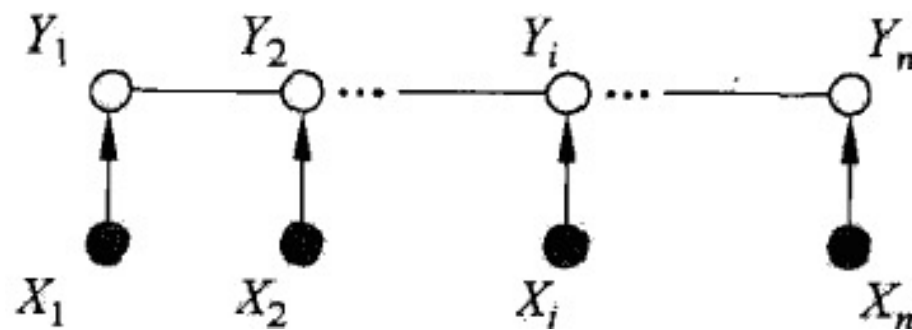


图 11.4 线性链条件随机场



$X$  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  $X=(X_1, X_2, \dots, X_n)$ ,  $Y=(Y_1, Y_2, \dots, Y_n)$  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

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

$i=1, 2, \dots, n$  (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 $k_1$ 个转移特征， $k_2$ 个状态特征， $K=k_1+k_2$ ，记

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

## 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

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

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

- 条件随机场可表示为：

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

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

# 条件随机场的简化形式

- 若 $w$ 表示权值向量： $w = (w_1, w_2, \dots, w_K)^T$

- 以 $F(y, x)$ 表示全局特征向量，即

$$F(y, x) = (f_1(y, x), f_2(y, x), \dots, f_K(y, x))^T$$

- 条件随机场写成内积：

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

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

# 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 $Y_0 = \text{start}$ ,  $Y_{n+1} = \text{stop}$ ，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 $i=1,2,\dots,n+1$ ，定义一个 $m$ 阶矩阵( $m$ 是标记 $Y_i$ 取值的个数)

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

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

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



# 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:

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

- 条件概率 $P_w(y|x)$ :

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

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

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

# 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:  $P(Y_i = y_i | x)$ ,  $P(Y_{i-1} = y_{i-1}, Y_i = y_i | x)$
  - 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  $i=0,1,\dots,n+1$ ，定义前向向量  $\alpha_i(x)$

$$\alpha_0(y|x) = \begin{cases} 1, & y = \text{start} \\ 0, & \text{否则} \end{cases}$$

- 递推公式：

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

- 又可表示为： $\alpha_i^T(x) = \alpha_{i-1}^T(x) M_i(x)$
- 即表示在位置  $i$  的标记是  $y_i$ ，且到位置  $i$  的前部分标记序列的非规范化概率， $y_i$  可取的值  $m$  个。所以是  $m$  维列向量。  
 $\alpha_i(x)$

# 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 $i=0,1,\dots,n+1$ ，定义后向向量  $\beta_i(x)$

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

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

- 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 $y_i$ ，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 :

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

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

- 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：

$$\begin{aligned} E_{P(Y|X)}[f_k] &= \sum_y P(y|x) f_k(y, x) \\ &= \sum_{i=1}^{n+1} \sum_{y_{i-1} y_i} f_k(y_{i-1}, y_i, x, i) \frac{\alpha_{i-1}^T(y_{i-1} | x) M_i(y_{i-1}, y_i | x) \beta_i(y_i | x)}{Z(x)} \\ &\quad k = 1, 2, \dots, K \end{aligned}$$

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

# 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 $f_k$ 关于联合分布 $P(X,Y)$ 的数学期望是：

$$\begin{aligned} & E_{P(X,Y)}[f_k] \\ &= \sum_{x,y} P(x,y) \sum_{i=1}^{n+1} f_k(y_{i-1}, y_i, x, i) \\ &= \sum_x \tilde{P}(x) \sum_y P(y|x) \sum_{i=1}^{n+1} f_k(y_{i-1}, y_i, x, i) \\ &= \sum_x \tilde{P}(x) \sum_{i=1}^{n+1} \sum_{y_{i-1} y_i} f_k(y_{i-1}, y_i, x, i) \frac{\alpha_{i-1}^T(y_{i-1}|x) M_i(y_{i-1}, y_i|x) \beta_i(y_i|x)}{Z(x)} \end{aligned}$$

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

# 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： $w = (w_1, w_2, \dots, w_K)^T$

- 向量增量： $\delta = (\delta_1, \delta_2, \dots, \delta_K)^T$

- 更新向量： $w + \delta = (w_1 + \delta_1, w_2 + \delta_2, \dots, w_K + \delta_K)^T$

- 关于转移特征 $t_k$ 的更新方程：
$$E_{\tilde{P}}[t_k] = \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n+1} t_k(y_{i-1}, y_i, x, i)$$
$$= \sum_{x,y} \tilde{P}(x) P(y|x) \sum_{i=1}^{n+1} t_k(y_{i-1}, y_i, x, i) \exp(\delta_k T(x,y))$$
$$k = 1, 2, \dots, K_1$$

# 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：

$$\begin{aligned} E_{\tilde{P}}[s_l] &= \sum_{x,y} \tilde{P}(x,y) \sum_{i=1}^{n+1} s_l(y_i, x, i) \\ &= \sum_{x,y} \tilde{P}(x) P(y|x) \sum_{i=1}^n s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x,y)) \\ l &= 1, 2, \dots, K_2 \end{aligned}$$

- $T(x,y)$ 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  $t_1, t_2, \dots, t_{K_1}$ ，  $s_1, s_2, \dots, s_{K_2}$ ；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  $k \in \{1, 2, \dots, K\}$ ，取初值  $w_k = 0$

(2) 对每一  $k \in \{1, 2, \dots, K\}$ ：

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

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

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  $k = K_1 + l$ ,  $l = 1, 2, \dots, K_2$  时, 令  $\delta_{K_1+l}$  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 $(x,y)$ 取值可能不同，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 $(x,y)$

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

# 条件随机场的学习算法

- 对于转移特征  $\delta_k$  的更新方程为：

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

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

- 其中：

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

# 条件随机场的学习算法

- 对于状态特征  $\delta_k$  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
- $T(x) = \max_y T(x, y)$  利用前向-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：

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

$a_{k,t}$  是特征  $t_k$  的期待值,  $\delta_k = \log \beta_k$ .  $\beta_k$  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：

$$\begin{aligned} E_{\tilde{P}}[s_l] &= \sum_{x,y} \tilde{P}(x) P(y|x) \sum_{i=1}^n s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x)) \\ &= \sum_x \tilde{P}(x) \sum_y P(y|x) \sum_{i=1}^n s_l(y_i, x, i) \exp(\delta_{K_1+l} T(x)) \\ &= \sum_x \tilde{P}(x) b_{l,t} \exp(\delta_k \cdot t) \\ &= \sum_{t=0}^{T_{\max}} b_{l,t} \gamma_l' \end{aligned}$$

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

# 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：

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



$$y^* = \arg \max_y P_w(y|x)$$

$$= \arg \max_y \frac{\exp(w \cdot F(y, x))}{Z_w(x)}$$

$$= \arg \max_y \exp(w \cdot F(y, x))$$

非规范化概率最大的最优路径



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

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

# 条件随机场的预测算法

- 路径表示标记序列：

$$w = (w_1, w_2, \dots, w_K)^T$$

$$F(y, x) = (f_1(y, x), f_2(y, x), \dots, f_K(y, x))^T$$

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

- 只计算非规范化概率：

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

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

- 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：

$$\delta_1(j) = w \cdot F_1(y_0 = \text{start}, y_1 = j, x), \quad j = 1, 2, \dots, m$$

- 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：

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

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

# 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：

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

- 及最优路径的终点： $y_n^* = \arg \max_{1 \leq j \leq m} \delta_n(j)$

- 由此最优路径终点返回：

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

- 得最优路径： $y^* = (y_1^*, y_2^*, \dots, y_n^*)^T$

# 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  $x = (x_1, x_2, \dots, x_n)$

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

(1) 初始化

$$\delta_1(j) = w \cdot F_1(y_0 = \text{start}, y_1 = j, x), \quad j = 1, 2, \dots, m$$

(2) 递推. 对  $i = 2, 3, \dots, n$

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

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

# 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

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

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

# • Q & A

7 七月在线  
JULYEDU.COM

## 从零学AI的路线图：挑战年薪40万

通过七月在线《机器学习集训营》可解锁全部技能 超过2000人的成功经验：从零学AI、传统IT转型AI

- 核心语法：循环 判断 控制
- 函数与面向对象
- 迭代器、生成器、装饰器

### Python基础

### 数据分析

- 科学计算之numpy
- 数据分析之pandas
- 数据分析实战（美国大选、房价预测）

### 数据结构

- 链表、队列、堆栈
- 字符串和数组
- 哈希表、树、图
- 查找与排序：增删改查
- 分治递归回溯
- 贪心和动态规划
- 概率与组合

### 大数据

- Hadoop 基础(HDFS与YARN)
- MapReduce与Hive SQL
- 分布式数据库Hbase
- Spark与Flink

### 机器学习

- 机器学习的基本流程
- 经典模型：线性模型、决策树
- 常考模型：SVM与XGBoost
- 应用模型：HMM与CRF
- 特征工程：数据处理、模型构建/调优
- 上线部署：模型调参与模型评估
- 项目实战：图像检索、金融风控

### 深度学习

- 核心模型：CNN RNN LSTM
- CV应用：Two-Stage和One-Stage框架
- 项目实战：调参、优化、模型压缩、蒸馏收敛
- 框架应用：TensorFlow与Pytorch

### CV NLP 推荐的企业级项目实战

- 大规模跨境追踪/重识别 (ReID)
- 人体关节点提取
- 智能客服系统
- 人体关节点提取
- 电商平台的商品推荐系统
- 从零开整电影推荐网站
- 简历指导、面试辅导、就业内推

#### CV 企业级项目实战

#### NLP 企业级项目实战

#### 推荐的企业级项目实战

阶段一 夯实AI基础：数据分析与数据结构

阶段二 掌握AI核心：大数据/ML/DL

阶段三 CV NLP 推荐项目实战