概率图 Inference

MAP 最大后验概率 $\hat{z}=\arg\max\limits_{z} \underbrace{P(z|X)}_{后验概率} \propto \arg\max\limits_{z} P(X,z)$

- 精确推断
  - 变量消除 Variable Elimination (VE)
  - 信念传播 Belief Propagation (BP) —— 采用VE思想 取长补短
    - e.g. Sum-Product Algorithm
  - 联合树 Junction Tree Algorithm —— BP在图结构的拓展
- 近似推断
  - Loop Belief Propagation (处理环结构)
  - Monte Carlo Inference
    - importance sampling
    - MCMC：马尔可夫链蒙特卡洛
  - variational inference 变分推断

一. Variable Elimination

$$a \to b \to c \to d \qquad P(d)=\sum_{a,b,c} P(a,b,c,d)$$

$$=\sum_{a,b,c} P(a)P(b|a)P(c|b)P(d|c) \quad (马氏链)$$

假设a,b,c,d均是离散的二值

即 $a,b,c,d \in \{0,1\}$.

$$P(d)=P(a=0)P(b=0|a=0)P(c=0|b=0)P(d\text{[redacted]}|c=0)+\cdots\cdots+$$
$$P(a=1)P(b=1|a=1)P(c=1|b=1)P(d\text{[redacted]}|c=1)$$

= [redacted] 8·因子项

简化 $P(d)=\sum\limits_{a,b,c} xxx \overset{\star}{=} \sum\limits_{b,c} P(c|b)P(d|c)\cdot \underbrace{\sum\limits_{a} P(a)P(b|a)}_{\phi_a(b)}$

$=\sum\limits_{c} P(d|c)$ [redacted] $\underbrace{\sum\limits_{b} P(c|b)\phi_a(b)}_{\phi_b(c)}$

$=\phi_c(d)$

利用 乘法对加法的分配律

VE缺点 ① computation order NP-Hard

② 重复计算

二、Belief Propagation（只适用于树结构）（Sum-Product）

VE复习：$a \xrightarrow{m_{a\to b}} b \xrightarrow{m_{b\to c}} c \xrightarrow{m_{c\to d}} d \xrightarrow{m_{d\to e}} e$

$$P(e)=\sum_{a,b,c,d} P(a,b,c,d,e)$$

$$=\sum_d P(e|d)\sum_c P(d|c)\sum_b P(c|b)\sum_a P(b|a)\,P(a)$$

$$=\sum_d P(e|d)\sum_c P(d|c)\sum_b P(c|b)\cdot \underline{m_{a\to b}(b)}$$

对应上节的 $\phi_a(b)$，是一种功能函数

$$=\sum_d \times\times \sum_c \times\times \cdot m_{b\to c}(c)$$

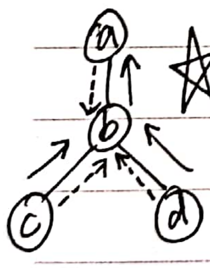
$$=\cdots$$

$$=m_{d\to e}(e)$$

$$P(c)=\sum_{a,b,d,e} P(a,b,c,d,e)$$

$$=\underbrace{\left(\sum_b P(c|b)\sum_a P(b|a)\,P(a)\right)}_{m_{b\to c}(c)\ \text{前向}}\cdot\underbrace{\left(\sum_d P(d|c)\sum_e P(e|d)\right)}_{m_{d\to c}(c)\ \text{反向}}$$

这是VE的drawback：重复计算



☆ $P(a,b,c,d)=\frac{1}{Z}\underbrace{\psi_a(a)\cdot\psi_b(b)\cdot\psi_c(c)\cdot\psi_d(d)\cdot\psi_{a,b}(a,b)\cdot\psi_{b,c}(b,c)\cdot\psi_{b,d}(b,d)}_{\text{4个点3条边}}$

$$P(a)=\sum_{b,c,d} P(a,b,c,d)$$

$$P(b)=\sum_{a,c,d} P(a,b,c,d)$$

$$\begin{cases} P(a)=\psi_a\, m_{b\to a}(x_a) \\ m_{b\to a}(x_a)=\sum_{x_b}\psi_b\,\psi_{a,b}\, m_{c\to b}(x_b)\, m_{d\to b}(x_b) \\ \qquad=\sum_{x_b}\psi_a\,\psi_{a,b}\prod_{k\in \underbrace{N(b)}_{\text{结点b的邻居}}-a} m_{k\to b}(x_b) \end{cases}$$

如何计算边缘概率?

$$P(x_i) = \psi_i \prod_{k\in NB(i)} m_{k\to i}(x_i)$$

$k\in NB(i)$, NB is "neighbor"

$$m_{j\to i}(x_i) = \sum_{x_j} \underbrace{\psi_{ij}}_{\text{边的势能函数}} \underbrace{\psi_j}_{\text{点的势能函数}} \prod_{k\in NB(i)-j} m_{k\to j}(x_j)$$

为了避免VE中的"重复计算"问题,只需要求所有的 $m_{i\to j}$ 即可计算所有的边缘概率

而对于BP算法: BP = VE + caching = m_{ij} 导出边缘概率 = 图的遍历

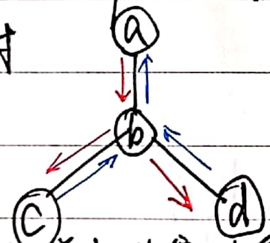
$m_{b\to a} = \sum_b \psi_{ab}\cdot belief(b)$, 其中 $belief(b) = \psi_b \cdot children$

对于 belief(b): 从b出发一共有多少信息量

$m_{b\to a}$: 从b到a的路径上能送给a的信息量

BP (Sequential Implementation)

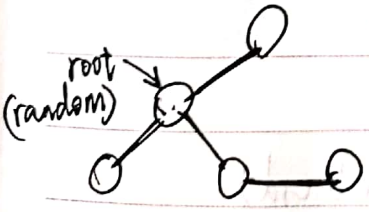
对于一棵无向树



1. get root (假设a是根结点)
2. collect msg (递归) 蓝线
3. distribute msg (递归) 红线

Collect Msg: /* 递归收集消息 */
for x_i in NB(root)
  collect Msg(x_i)

BP (Parallel Implementation)



先收集已经来到该结点的消息

把目前保存的信息分发给全部结点.

这个并行的算法会使整个系统收敛.

三. Max-Product Algorithm

Graph $=\{X, E\}$

(I) 边缘概率 $E=\{e_1,\cdots,e_k\}$ $P(E)$ likelihood

(II) 条件概率 $P(X|E)$ posterior $\longrightarrow$ $X=(Y,Z)$ 多元随机变量

$P(Y|E)=\sum_{Z} P(X|E)$

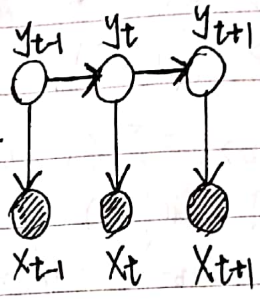
(四) MAP $\hat{x}=\arg\max_{X} P(X|E)$

$\hat{Y}=\arg\max_{Y} P(Y|E)$

$=\arg\max_{Y} \sum_{Z} P(X|E)$

[redacted]

对于HMM: Decoding: $\hat{Y}=\arg\max_{Y} P(Y|X)$ $\longleftarrow$



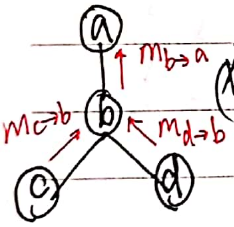
↳ Viterbi

↳ 动态规划

而对于 Max-Product:

① 是 BP 的改进

② 是 Viterbi 的泛化



$(x_a^*, x_b^*, x_c^*, x_d^*)=\arg\max_{x_a x_b x_c x_d} P(x_a,x_b,x_c,x_d|E)$

E: evidence

$m_{b\to a}$: 能让 $P(x_b,x_c,x_d|E)$ 达到 max 的值

$m_{c\to b}$: 经过 c 结点能让 $P(x_c|E)$ 达到 max 的值

$$m_{j\to i}=\max_{x_j} \psi_j \psi_{ij} \prod_{k\in NB(j)-i} m_{k\to j}$$

Note: 对于 BP: $\max_{x_j} \longrightarrow \sum_{x_j}$

e.g. $m_{c\to b}=\max_{x_c} \underbrace{\psi_c \psi_{bc}}_{\text{对c的一个函数}}$

$\max P(a,b,c,d)$

$=\max_{a} \underbrace{\psi_a \cdot m_{b\to a}}_{\text{关于a的函数}}$

$m_{d\to b}=\max_{x_d} \psi_d \psi_{bd}$

$m_{b\to a}=\max_{x_b} \underbrace{\psi_b \psi_{ab}\, m_{c\to b}\, m_{d\to b}}_{\text{b的函数}}$

精确推断总结:

1、Max-Product是一种动态规划，要求每一个子问题都能达到概率最大

2、Sum-Product与Max-Product的区别:

都是图的精确推断

max-product就是sum-product把求和符号换成max

3，边缘概率

$$P(X_i)=\psi_i\prod_{k\in NB(i)} m_{j\to i}(X_i)$$

$$m_{j\to i}(X_i)=\sum_{X_j}\psi_{ij}\psi_j\prod_{k\in NB(j)-i} m_{k\to j}(X_j)$$