概率图模型 — 背景介绍 (1)

正

把数据结构里增加图进去。 图是用于表达的工具

数据往往是一个高维的随机变量 $p(x_1, x_2, \cdots x_p)$

$$\begin{cases} 边缘概率\ p(x_i) \\ 条件概率\ p(x_j|x_i) \end{cases}$$

$$\begin{cases} \text{Sum 求和}\quad p(x_1) = \int p(x_1, x_2)\,dx_2 \rightarrow 边缘概率 \\ \text{求积}\quad p(x_1, x_2) = p(x_1)\cdot p(x_2|x_1) = p(x_2)\cdot p(x_1|x_2) \end{cases}$$

Chain Rule : $p(x_1, x_2, \cdots x_p) = \prod_{i=1}^{p} p(x_i|x_1, x_2 \cdots x_{i-1})$

Bayesion Rule : $p(x_2|x_1) = \frac{p(x_1, x_2)}{p(x_1)} = \frac{p(x_1, x_2)}{\int p(x_1, x_2)\,dx_2} = \frac{p(x_2)\cdot p(x_1|x_2)}{\int p(x_1, x_2)\,dx_2}$

困境：维度高，计算复杂. Chain Rule $p(x_1 x_2 \cdots x_p)$ 计算量太大 ▲

简化：① 相互独立（维度之间）. 那么 $p(x_1 x_2 \cdots x_p) = \prod_{i=1}^{p} p(x_i)$，朴素贝叶斯 $\rightarrow p(x|y) = \prod_{i=1}^{p} p(x_i|y)$

但是相互独立太强了，使用马尔可夫（一阶） $x_j \perp x_{i+1} | x_i,\ j<i \longrightarrow \boxed{条件独立性}$ ▲ $x_A \perp x_B | x_C$ （x_A, x_B, x_C 是集合，且不相交）

- 概率图
  - Representation 表示
    - 有向图 Bayesian Network
    - 高斯图（连续）
      - Gaussian BN
      - Gaussian MN
    - 无向图 Markov Network
  - Inference 推断
    - 精确
    - 近似
      - 确定性（VI：变分）
      - 随机（MCMC）
  - Learning 学习
    - 参数学习
      - 完备数据
      - 隐变量 → EM
    - 结构学习（图结构）

概率图模型 — 贝叶斯网络 — 条件独立性 (2)

如何构建有向图？拓扑排序. 若 $p(x_i|x_j)$，那么 ⓧⱼ → ⓧᵢ，x_j 是 x_i 的父节点

① $b \leftarrow a \rightarrow c$

$$\left.\begin{array}{l} p(a,b,c) = p(a)\cdot p(b|a)\cdot p(c|a) \rightarrow 看图因式分解 \\ p(a,b,c) = p(a)\cdot p(b|a)\cdot p(c|a,b) \rightarrow \text{Chain Rule} \\ c \perp b | a \therefore p(c|a) = p(c|a,b) \Rightarrow c \perp b | a\ 条件独立 \end{array}\right\}\ 证明了图与概率独立性$$

若 a 被观测，则路径被阻塞，$c \perp b$.

② ⓐ → ⓑ → ⓒ

若 b 被观测，则路径被阻塞，$a \perp c$

③ $a \rightarrow c \leftarrow b$

$p(a,b,c) = p(a)\cdot p(b)\cdot p(c|a,b)$

$\Rightarrow p(b) = p(b|a) \Rightarrow a \perp b$

默认 $a \perp b$，路径阻塞

若 c 被观测，路径是通的

可参考亲戚之间关系

①: tail to tail

②: head to tail

③: head to head

白板（9）

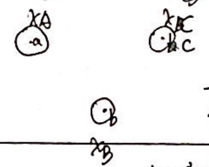
概率图模型—贝叶斯网络—D划分（3）

反

条件独立性 $x_A \perp x_C \perp x_B$

D-Seperation：对以上三种情况更适用

方法的描述 $\begin{cases} a,c \text{ 若满足①,②，则 } b \text{ 一定在 } x_b \text{ 里} \\ \text{……③……不在 } x_b \text{ 里，且其后续节点都不会在 } x_b \text{ 里} \end{cases}$

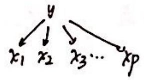


→ 全局 Markov property（全局马尔可夫性）

→ 若 A, B, C，$sep(A,C|B)$，则 $x_A \perp x_C | x_B$

概率图模型—贝叶斯网络—例子（4）

Naive Bayes



$x_1 \perp x_2 | y$

Bayesian Network

从单一到混合

从有限到无限

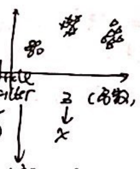
空间（离→连）　时间

- 单一：Naive Bayes（→分类）→ $p(x|y) = \prod_{i=1}^{p} p(x_i|y=1)$（→p维）
- 混合：GMM → 聚类
- 时间：
  - Markov Chain
  - Gassian process（高斯过程，无限维高斯分布）
- 连续：Gaussian Network，Gaussan Bayness Network（有向）

→ 动态模型

HMM（离散z）　LDS → Kalman Filter（连续线性）　particle Filter

GMM



z（离散，$z=1,2\dots k$）

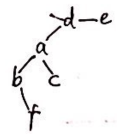
$z \to x$

非高斯（连续），非线性.

概率图模型—马尔可夫随机场（5）

① 全局马尔可夫. $x_A \perp x_C | x_B$

② 局部马尔可夫　$a \perp \{全-a-a邻居\} | 邻居$



$a \perp \{e,f\} | \{b,c,d\}$

③ 成对马尔可夫　$x_i \perp x_j | x_{-ij}$　$(i \neq j)$

除去i与j

以上是条件独立性体现的3个方面.

①⟺②⟺③，3者等价，可互推出

团：一个关于结点的集合，集合的结点之间相互联通

最大团：无法再往内加入任何一个结点.

无向图的因子分解.

$$p(x) = \frac{1}{z}\prod_{i=1}^{K}\phi(x_{c_i})$$

x_{c_i}：一个团对应随机变量的集合.

z：为一化因子

$$z = \sum_{x}\prod_{i=1}^{K}\phi(x_{c_i}) = \sum_{x_1}\cdots\sum_{x_p}\prod_{i=1}^{K}\phi(x_{c_i})$$

团 vs. 条件独立性.

如何对应？

???

概率图模型 — 马尔可夫随机场 (6)

(正)

Markov
① 全局
② 局部
③ 成对

① ⟺ ② ⟺ ③

因子分解

$p(x)=\frac{1}{Z}\prod_{i=1}^{k}\psi(x_{c_i})$

c_i：最大团　x_{c_i}：最大团的随机变量集合

$\psi(c_{x_i})$：势函数，必为正

$Z=\sum_{x}\prod_{i=1}^{k}\psi(x_{c_i})=\sum_{x_1}\sum_{x_2}\cdots\sum_{x_p}\prod_{i=1}^{k}\psi(x_{c_i})$

$\psi(x_{c_i})=\exp\{-E(x_{c_i})\}>0$

↳ energy function

→ $p(x)$ 称为 Gibbs Distribution (Boltzmann Distribution)

统计物理学知识

① ⟺ ② ⟺ ③ ⟺ 因子分解（基于最大团）

此为 Hammesley-Clifford 定理 ▲ — ①定理

最大熵原理 ⟹ 指数族分布（也叫 Gibbs 分布）$\left(p(x)=h(x)\cdot\exp\{\eta^T\phi(x)-A(\eta)\}=\frac{1}{Z(\eta)}\cdot h(x)\cdot\exp\{\eta^T\phi(x)\}\right)$

而 $p(x)=\frac{1}{Z}\prod_{i=1}^{k}\psi(x_{c_i})=\frac{1}{Z}\prod_{i=1}^{k}\exp\{-E(x_{c_i})\}=\frac{1}{Z}\exp\left\{-\sum_{i=1}^{k}E(x_i)\right\}$ 指数族分布

Markov Random Field $\underset{\text{等价}}{\Longleftrightarrow}$ Gibbs Distribution ⟸ 最大熵原理 [通过①定理]

概率图模型 — Inference (7)

Inference → 求概率

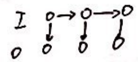
联合概率 $p(x)=p(x_1,x_2\cdots x_p)$

边缘概率 $p(x_i)=\sum_{x_1}\sum_{x_2}\cdots\sum_{x_{i-1}}\sum_{x_{i+1}}\cdots\sum_{x_p}p(x)$
（积分掉）

条件概率 $p(x_A|x_B)$，$x=x_A\cup x_B$
（已知部分，求另一部分）

MAP: $\hat{z}=\underset{z}{\arg\max}\,p(z|x)\propto\underset{z}{\arg\max}\,p(z,x)$ 等价于求联合概率

$\left(p(z|x)=\frac{p(x,z)}{p(x)}\right)$



HMM 三大问题 → 边缘概率

① Evaluation: $p(O)=\sum_{I}p(I,O)$ ⟹ { 前向 / 后向

② Learning: $\hat{\lambda}$

③ Decoding: $\hat{I}=\underset{I}{\arg\max}\,p(I|O)$ → Viterbi (DP)

①③都是 Inference 问题

HMM：特殊的 Dynamic Bayies Network

Inference
- 精确
  - variable Elimanatim (VE)
  - Belief propagation (BP) → Sum-product Algorithm [针对树结构]
  - Junction Tree Algorithm ← 普通图结构
- 近似
  - Loop Belief propagation (有环图)
  - Monte Carlo Inference: Important Sampling, MCMC
  - Variational Inference

白板(9)　　概率图模型 - variable elimination (8)　　(反)

(Inference 问题) 精确

变量消除法 variable elimination ⟶ 乘法分配律

Tasks: Inference (给定 $p(x)=p(x_1,x_2\cdots x_p)$)

- 边缘概率 $p(x_1)=\sum_{x_1x_2\cdots x_p} p(x_1,x_2\cdots x_p)$
- 条件概率 $p(x_A|x_B)$
- MAP $\tilde{x}_A=\underset{x_A}{\arg\max}\, p(x_A|x_B)=\arg\max p(x_A,x_B)$

$a\to b\to c\to d$? (假设 a,b,c,d 是二值变量 $a,b,c,d\in\{0,1\}$)

$p(d)=\sum_{a,b,c} p(a,b,c,d)$　　因子分解　　复杂度 k^p $(k=2,\ p=4)$

$=\sum_{a,b,c} p(a)\cdot p(b|a)\cdot p(c|b)\cdot p(d|c)$

$=p(a=0)\cdot p(b=0|a=0)\cdot p(c=0|b=0)\cdot p(d=0|c=0)$

$+p(a=1)\cdot p(b=0|a=1)\cdot p(c=0|b=0)\cdot p(d=0|c=0)$

$+\ \vdots$

$+p(a=1)\cdot p(b=1|a=1)\cdot p(c=1|b=1)\cdot p(d=1|c=1)$

$=16$ 个 · 因子积

换顺序

$=\sum_{b,c} p(c|b)\cdot p(d|c)\cdot \underbrace{\sum_a p(a)\cdot p(b|a)}$ $\to =\sum_a p(a,b)=p(b)$ 写作 $\phi_a(b)$：只与 b 相关的函数，从 a 过来的

$=\sum_c p(d|c)\cdot \underbrace{\sum_b p(c|b)\cdot \phi_a(b)}_{\phi_b(c)}$

$=\phi_c(d)$ ⟶ 只与 d 相关的函数

(乘法对加法的分配律) → 核心思想 $ab+ac=a(b+c)$

缺点：① 没有把计算存储下来。如求 d 算一次，求 c 算一次，重复计算。

② 消除次序 ordering → NP-hard

找到最优次序 NP-hard

适用于节点少，结构少，但是是后序的基础。

Belief propagation (9)

解决 VE 重复计算问题。

组合起来(先求结果)

$a\xrightarrow{m_{a\to b}(b)} b\xrightarrow{m_{b\to c}(c)} c\xrightarrow{m_{c\to d}(d)} d\xrightarrow{m_{d\to e}(e)} e$ [类似 HMM F/B 算法]

$p(a,b,c,d,e)=p(a)\cdot p(b|a)\cdot p(c|b)\cdot p(d|c)\cdot p(e|d)$

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

$=\cancel{\sum_{a,b,c,d}}\ \underbrace{\sum_d p(e|d)\cdot \underbrace{\sum_c p(d|c)\cdot \underbrace{\sum_b p(c|b)\cdot \underbrace{\sum_a p(b|a)\cdot p(a)}_{m_{a\to b}(b)}}_{m_{b\to c}(c)}}_{m_{c\to d}(d)}}_{m_{d\to e}(e)}$ 省略

若再求 $p(c)$

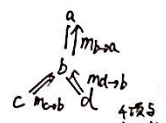
$p(c)=\sum_{a,b,d,e} p(a,b,c,d,e)$

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

$m_{b\to c}(c)$ → 重复计算

F-B 算法(类似 HMM)

Chain → Tree　有向 → 无向 [无向树]



4项与 b 相关

$p(a,b,c,d)$

$=\frac{1}{Z}\psi_a(a)\cdot\psi_b(b)\cdot\psi_c(c)$

$\cdot\psi_d(d)\cdot\psi_{ab}(ab)\cdot\psi_{bc}(bc)$

$\cdot\psi_{bd}(b,d)$

[4个点，3个边]

从信息角度考虑

$\underbrace{\psi_a\ \underbrace{\sum_b\left[\underbrace{\sum_c \psi_c\cdot\psi_{bc}}_{m_{c\to b}(b)}\ \psi_b\ \underbrace{\sum_d \psi_d\cdot\psi_{bd}}_{m_{d\to b}(b)}\ \psi_{ab}\right]}_{m_{b\to a}(a)}}_{p(a)}$

$m_{b\to a}(x_a)=\sum_{x_b}\psi_{ab}\cdot\psi_b\, m_{c\to b}(x_b)\cdot m_{d\to b}(x_b)$ 与 b 相关，4项

$p(a)=\psi_a\, m_{b\to a}(x_a)$

通识写法

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

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

这就是 VE 在无向树表达

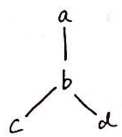
NB 代表 nerber，邻居

结论：不要直接求边缘概率 $(p(a),p(b),p(c)\cdots)$

只需求 $m_{i\to j}$，只求 6 个 $m_{i\to j}$ 就够存起来 ⟶ BP 算法 (本例)

Belief propagation

Belief propagation (10)



$m_{j\to i}(x_i)=\sum_{x_j}\psi_{ij}\,\psi_j\prod_{k\in NB(j)-i} m_{k\to j}(x_j)$ —— ①

$p(x_i)=\psi_i\prod_{k\in NB(i)} m_{k\to i}(x_i)$ —— ②

对于①式 $m_{j\to i}=\sum_j \psi_{ij}\,\psi_j\prod_{\text{子节点}} m_{k\to j}$，以左图为例 $m_{b\to a}=\sum_b \psi_{ab}\cdot\overbrace{\boxed{\underbrace{\psi_b}_{\text{self}}\cdot\underbrace{m_{c\to b}\,m_{d\to b}}_{\text{children}}}}^{\text{belief}(b)}$

$\therefore\begin{cases}\text{belief}(b)=\psi_b\cdot\text{Children}\\ m_{b\to a}=\sum_b\psi_{ab}\cdot\text{belief}(b)\end{cases}$

分两步，可从信息汇总角度去理解.
belief(b) 代表 b 点的总信息量. $m_{b\to a}$ 就是从 $b\to a$ 的信息量 + belief(b) 本身信息量

正是因为在求解过程中. m_{ij} 多计算. ∴才有 BP 算法

BP = VE + Caching (缓存)
- → 直接求 $m_{ij}\Rightarrow p(x_i)$
- → 树的遍历

BP (Sequential Implementation)

① get root → ~~都~~ 节点都可以. a, c, d

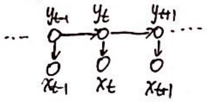
② Collect message
for x_i in NB(root):
  collect message(x_i)

③ Distribute message
for x_j in NB(root):
  distribute message(x_j)

可得 m_{ij} for all 的 i,j ~~∈~~ V → 节点
从而 $p(x_k)$, $k\in V$

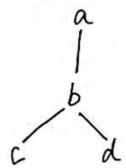
Max product (11)(12)

HMM 的 Decoding 问题
求隐状态序列最大



Decoding: $\hat{Y}=\arg\max_Y P(Y|X)$

Viterbi (动态规划)



$(\hat a,\hat b,\hat c,\hat d)=\arg\max_{a,b,c,d} p(x_a,x_b,x_c,x_d|E)$

E是观测的其它节点,

Belief propagation (树结构)

~~$m_{j\to i}(x_i)=\sum_{x_j}\psi_j(x_j)\psi_{ij}(x_i,x_j)\cdot\prod_{k\in NB(j)-i} m_{k\to j}(x_j)$~~

Max product

① BP 的改进 ② Viterbi 的推广

$m_{j\to i}=\max_{x_j}\psi_j\cdot\psi_{ij}\prod_{k\in NB(j)-i} m_{k\to j}$ [把之前 $\sum_{x_j}$ 用 $\max_{x_j}$ 代替]

对里有k个状态, 找最大状态 ▲
[之前是对所有状态求和]

之前 Sum-product，现在 Max-product. 其它不变.

一环套一环，Viterbi 思想

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

$m_{c\to b}=\max_{x_c}\underbrace{\psi_c\cdot\psi_{b,c}}_{\text{关于c的函数}}$

$m_{d\to b}=\max_{x_d}\underbrace{\psi_d\cdot\psi_{bd}}_{\text{关于d的函数}}$

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

$\max p(a,b,c,d)=\max_{x_a}\underbrace{\psi_a\cdot m_{b\to a}}_{\text{关于a的函数}}$
$=\max p(x_a,x_b,x_c,x_d)$

$x_a^*, x_b^*, x_c^*, x_d^*$

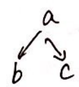
Factor Graph 因子图 (13)　(反)

BP只应用于树结构，引入环后，使用因子图.

有向图：$p(x)=\prod_{x} p(x_i|x_{pa(i)})$

无向图：$p(x)=\frac{1}{Z}\prod_{i=1}^{K}\phi_{ci}(x_{ci})$ ↳最大团集合

道德图. 有向图 ⟶ 无向图 (tree-like graph)
(树)　　　(引入环)



$p(a,b,c)=p(a)\cdot p(b|a)\cdot p(c|a)$

$p(a,b,c)=\frac{1}{Z}\phi_1(a,b)\cdot\phi_2(a,c)$

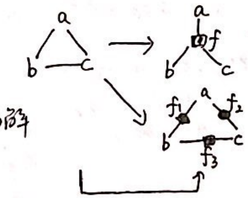
因子图 $x=x_1,x_2\cdots x_p$

$p(x)=\prod_{s} f_s(x_s)$　s：图的节点子集
x_s：x的随机变量子集

② 简便

因式分解本身对应一个特殊的因子图

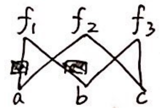
因子图：看作对因式分解的更进一步分解



有环 ⟶ 无环

$p(x)=f(a,b,c)$　因子图不唯一

$p(x)=f_1(a,b)+f_2(a,c)+f_3(b,c)$

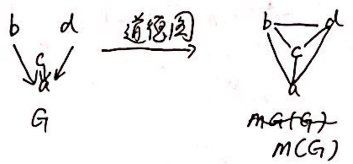


Moral Graph 道德图 (14)

产生原因：想把有向图转为无向图.
因为更为通用 [令归一化因子 Z=1]

有向图：head-head (V结构) — 特殊 ▲

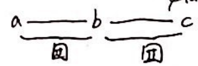
处理办法：
① $\forall x_i\in G$，将 parent(x_i) 两两相连 (↗父集合)
② 将G中有向边替换成无向边



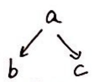
Sep(A,B|C) ⟺ D-Sep(A,B|C)
道德图 ↗　　不证明了.

① a ⟶ b ⟶ c

$p(a,b,c)=p(a)\cdot\underbrace{p(b|a)}_{\phi(a,b)}\cdot\underbrace{p(c|b)}_{\phi(b,c)}$

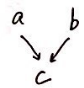


②

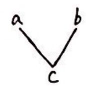


$p(a,b,c)=p(a)\underbrace{p(b|a)}_{\phi(a,b)}\cdot\underbrace{p(c|b)}_{\phi(b,c)}$

③



$p(a,b,c)=\underbrace{p(a)\cdot p(b)\cdot p(c|a,b)}_{\phi(a,b,c)}$



(a, b, c 显然不是一个团 → ∵ a,b 不通)

这是团!