Probability Graph 05 Markov Network

Chen Gong

27 November 2019

上一小节中，我们分析了有向图 Bayesian Network，得到了因子分解法，$p(x) = \prod_{i=1}^{N} p(x_i|x_{pa(i)})$。虽然,有向图中可以方便直观的表达条件独立性,但是它也有它的局限性。也就是我们提到的对于 Head to Head 的结构来说，当中间节点被观察到的时候，反而是两端的节点是相关的。这违反了条件独立性的特点，也就是当某些变量被观察到时，其他变量之间是独立的特点，这种情况有点反常，并不太好办。

但是，在无向图中就完全不会出现这样的情况，因为本来就没有方向，而且在无向图中也有类似的 D-Separation 性质。

1 Condition Independence in Markov Network

Markov 中的条件独立，大致可以被我们分成三种情况，Global Markov，Local Markov 和 Pair Markov。

1.1 Global Markov

假设现在有三个集合 $X_A \perp X_B | X_C$，我们想得到 $a \in X_A, b \in X_B$ 之间相互独立，这个应该怎么办？我们给出，只有 a 和 b 的中间节点至少有一个位于 c 中，那么我们就可以得到 $a \perp c$。

1.2 Local Markov

我们以下图的一个 Markov Network 为例，

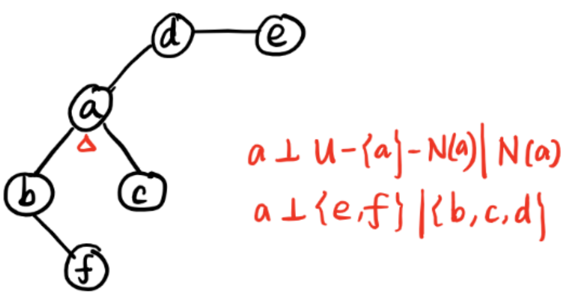

图 1: Markov Network 示意图

用文字的语言来描述就为 $a \perp \{$ 全集 $-a-$ 邻居 $\}|$ 邻居。那么在这个图中，我们就可以表示为 $a \perp \{e, f\}|\{b, c, d\}$。

1.3 Pair Markov

成对马尔可夫性质可以被我们描述为：$x_i \perp x_j | x_{-i-j}$ $(i \neq j)$。其中，x_{-i-j} 为从全集中去掉 x_i 和 x_j 而留下了的集合。

那么条件独立性就可以提现在，Global，Local 和 Pair 中。其中 Global⇔Local⇔Pair。也就是这三种条件独立的方法得到的结果是一样的。

2 因子分解法

我们想一想在一个无向图中，如何来体现我们想要的条件独立性。这里的引入和之前的不太一样，我们首先需要引入几个概念。**团**: 这是一个关于节点的集合，集合中的节点是相互连通的。而最大团，就很好理解了吧，也就是最大的连通子集。我们可以将无向图的分离定义到团上，我们假设 $c_1, c_2, \cdots, c_k$ 表示为团。那么，我们可以将联合概率定义为：

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

其中，z 是归一化因子，因为没有归一化因子的话，这不能被称为一个概率分布，因为概率分布首先就要保证和等于 1。那么，z 被定义为

$$z = \sum_{x} \prod_{i=1}^{k} \phi(x_{c_i}) = \sum_{x_1} \cdots \sum_{x_N} \prod_{i=1}^{k} \phi(x_{c_i}) \tag{2}$$

而这里的因子分解法必定和 Global，Local 和 Pair 都是等价的，但是现在根本就看不出来，那么我们看看后面会怎么说，哈哈哈！

3 Gibbs Distribution 和 Exponential Distribution

这个部分是上面部分的一个加深理解，首先我们需要总结一下。

1. Global Markov：$X_A \perp X_C | X_B$。如果 A, B, C, 满足 $Sep(A, C|B)$，Sep 代表 D-Separation，那么 $X_A \perp X_C | X_B$。

2. Local Markov Network：$x_i \perp x_{-i-nb(i)} | x_{nb(i)}$，其实 nb(i)：neighbor of node i。

3. Pair Markov：$x_i \perp x_j | x_{-i-j}$。

1⇔2⇔3⇔ 因子分解 (基于最大团)。这里面实际上是有个 Hammesley-Clifford 定理的，这个定理的证明非常的困难，我们这里就不做过多的阐述 (实际上我也看的有点懵逼，有需求的小伙伴可以查下 Hammesley-Clifford 定理的证明)。

因子分解：

在我们之前的定义中，

$$p(x) = \frac{1}{Z} \prod_{i=1}^{k} \phi(x_{c_i}) \tag{3}$$

c_i：最大团；x_{c_i}：最大团的随机变量集合；$\phi(x_{c_i})$：势函数，必须为正。这里的概念都是来自于统计物理学和热力学的过程。这里的势函数还有可以做文章的地方。

因为，势函数必定为正，我们可以将势函数表达为 $\phi(x_{c_i}) = \exp\{-E(x_{c_i})\}$。其中，$E(x_{c_i})$ 称为 Energy function。实际上用这种形式表达的 $p(x)$，为 Gibbs Distribution，或者又被称之为 Boltzman Distribution。有了 $\phi(x_{c_i})$ 的形式，我们可以进一步推导得：

$$p(x) = \frac{1}{Z}\prod_{i=1}^{K}\phi(x_{c_i}) = \frac{1}{Z}\prod_{i=1}^{K}\exp\{-E(x_{c_i})\} = \frac{1}{Z}\exp\{-\sum_{i=1}^{k}E(x_{c_i})\} \tag{4}$$

我们再来回顾一下指数族分布，指数族分布的通用表达形式为：

$$p(x) = h(x)\exp\{\eta^T\phi(x) - A(\eta)\} = h(x)\frac{1}{Z(\eta)}\exp\{\eta^T\phi(x)\} \tag{5}$$

在这里我们把 $\exp\{-A(\eta)\}$，直接记为 $Z(\eta)$。大家观察就会发现势函数也就是 Gibbs Distribution 就是一个指数族分布。Gibbs 是来自统计物理学，形式上和指数族分布时一样的。而指数族分布实际上是由最大熵分布得到的，那么我们可以理解 Gibbs 分布也是有最大熵原理得到的。而马尔可夫随机场 (Markov Random Field) 实际上等价于 Gibbs 分布。至于为什么？这实际上全部都在 Hammesley-Clifford 定理中，有兴趣的同学，请自行查阅。