

Proof of Hammersley-Clifford Theorem

[TOC]

最近看语义分割论文DeepLab, 有使用全连接CRF恢复局部的细节信息, 提升分割精度。又回去复习了下CRF, 仍然有一个问题很困扰: "根据Hammersley Clifford定理, 一个无向图模型的概率可以表示为定义在图上所有最大团上的势函数的乘积"; 为什么可以这么定义, 也就是Hammersley Clifford定理证明过程, 书中并没有给出; 网上看到也有一些童鞋有同样的困惑, 本文翻译并备注了证明过程, 希望对大家有所帮助。

原文地址: [Proof of Hammersley-Clifford Theorem](#)

依赖知识

a): 熟悉概率论的基础知识

b): 了解概率图模型, 熟悉MRF, 最大团相关知识

定义1

一个无向图模型G称之为马尔科夫随机场(MRF), 如果两个顶点被观测顶点分割情况下条件独立。也就是说对图中任意顶点 X_i , 以下条件属性成立

$$P(X_i | X_{G \setminus i}) = P(X_i | X_{N_i}) \quad (1)$$

$X_{G \setminus i}$ 代表除了 X_i 之外的所有顶点, X_{N_i} 代表 i 的所有邻居顶点-即所有与 X_i 相连的顶点。

定义2

在无向图模型G上的一个概率分布 $P(X)$ 称之为吉布斯分布, 如果它能够因子分解为定义在团(clique)上的正函数的乘积, 这些团覆盖了G的所有顶点和边。即

$$P(X) = \frac{1}{Z} \prod_{c \in C_G} \phi_c(X_c) \quad (2)$$

C_G 是G上所有(最大)团的集合, $Z = \sum_x \prod_{c \in C_G} \phi_c(X_c)$ 是归一化常量。

证明过程

Hammersley Clifford告诉我们这两个定义是等价的, 下面将证明这个定理。

反向证明(吉布斯分布=>MRF)

设 $D_i = N_i \cup X_i$ 是包含 X_i 邻居顶点和 X_i 本身的集合。从等式(1)的右边开始

$$P(X_i | X_{N_i}) = \frac{P(X_i, X_{N_i})}{P(X_{N_i})} \quad (3)$$

$$= \frac{\sum_{G \setminus D_i} \prod_{c \in C_G} \phi_c(X_c)}{\sum_{x_i} \sum_{G \setminus D_i} \prod_{c \in C_G} \phi_c(X_c)} \quad (4)$$

基于是否包含 X_i 将最大团 C_G 分为两组:

$C_i = c \in C_G : X_i \in c$ 和 $R_i = c \in C_G : X_i \notin c$; 现在可以将等式(4)分为 C_i 和 R_i 上的乘积。

$$P(X_i | X_{N_i}) = \frac{\sum_{G \setminus D_i} \prod_{c \in C_i} \phi_c(X_c) \prod_{c \in R_i} \phi_c(X_c)}{\sum_{x_i} \sum_{G \setminus D_i} \prod_{c \in C_i} \phi_c(X_c) \prod_{c \in R_i} \phi_c(X_c)} \quad (5)$$

$$= \frac{\prod_{c \in C_i} \phi_c(X_c) \sum_{G \setminus D_i} \prod_{c \in R_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c \in C_i} \phi_c(X_c) \sum_{G \setminus D_i} \prod_{c \in R_i} \phi_c(X_c)} \quad (5)$$

在 $G \setminus D_i$ 上的求和可以移到 C_i 乘积的后面, 因为 C_i 团中所有的顶点一定都来自 D_i ; 因为 C_i 只包含 X_i 和与 X_i 相邻的顶点, 由 D_i 的定义可知; 因而 C_i 乘积对于在 $G \setminus D_i$ 上的求和相当于常数项, 故可以把 C_i 乘积拿到 $G \setminus D_i$ 上的求和的外面。

同样注意到因子 $\sum_{G \setminus D_i} \prod_{c \in R_i} \phi_c(X_c)$ 没有包含 X_i ,并且可以从分母移除, 因为分子也包含了它。因此有:

$$P(X_i | X_{N_i}) = \frac{\prod_{c \in C_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c \in C_i} \phi_c(X_c)} \quad (7)$$

$$= \frac{\prod_{c \in C_i} \phi_c(X_c)}{\sum_{x_i} \prod_{c \in C_i} \phi_c(X_c)} * \frac{\prod_{c \in R_i} \phi_c(X_c)}{\prod_{c \in R_i} \phi_c(X_c)} \quad (8)$$

$$= \frac{\prod_{c \in C_G} \phi_c(X_c)}{\sum_{x_i} \prod_{c \in C_G} \phi_c(X_c)} \quad (9)$$

$$= \frac{P(X)}{P(X_{G \setminus i})} \quad (10)$$

$$= P(X_i | X_{G \setminus i}) \quad (11)$$

消除了 $G \setminus D_i$ 上的求和项后, 在公式(8)的分子分母乘上一个相同的因子, 再次引入势函数; 最终公式(11)与公式(1)的左边相等, 证明了反向等价。

正向证明(MRF=>吉布斯分布)

对于任意 $s \subset G$, 定义一个如下的候选势函数:

$$f_s(X_s = x_s) = \prod_{z \subset s} P(X_z = x_z, X_{G \setminus z} = 0)^{-1^{|s|-|z|}} \quad (12)$$

1. 等式右边的乘积是在 s 的所有子集上进行的。
2. 对于 s 任意子集 z , $P(X_z = x_z, X_{G \setminus z} = 0)$ 表示属于 z 的顶点(随机变量取值)与 s 一致, 图中其它顶点给默认值(记做"0")。
3. 当 s 集合与 z 集合顶点个数不同时指数为1, 否则为0; $|s|$ 表示集合 s 中元素(顶点)个数。
4. 很显然 f 是正函数, 概率都是非负的。
5. 只需要需要证明如下两点, 即可说明无向图模型的概率 $P(X)$ 可以表示为图上所有团的势函数乘积。
6. $\prod_{s \subset G} f_s(X_s) = P(X)$ (a)
7. $f_s(X_s) = 1$ 如果 s 不是一个团

证明第一点

为证明第一点, 先来展示一个恒等式:

$$0 = (1-1)^K = C_0^K - C_1^K + C_2^K - \dots + (-1)^K C_K^K \quad (13)$$

- a. C_N^K 表示从 K 个元素中选取 N 个元素的所有组合情况
- b. 现在证明 $\prod_{s \subset G} f_s(X_s)$ 中所有的因子都可以互相抵消, 除了 $P(X)$;
- c. 对于任意子集 $z \in G$, 及 z 相关的因子 $\Delta = P(X_z, X_{G \setminus z} = 0)$; 它在 s 不包含 z 的情况下没有出现(此时 z 不会是 s 的子集) ;
- d. 它在 $s = z$ 情况下出现一次($z=s$ 是 s 的子集), 因而 $\Delta^{-1^0} = \Delta$
- f. 它在 s 包含 z 以及另外一个元素的情况下出现 $C_1^{|G|-|z|}$ 次; 因为 s 的选择有 $C_1^{|G|-|z|}$ 种, 并且满足 $|s| - |z| = 1$, 因此这时 $\Delta^{-1^1} = \Delta^{-1}$ 。
- g. 它在 s 包含 z 以及另外两个元素的情况下出现 $C_2^{|G|-|z|}$ 次; 因为 s 的选择有 $C_2^{|G|-|z|}$ 种, 并且满足 $|s| - |z| = 2$, 因此这时 $\Delta^{-1^2} = \Delta$
- h. 依次类推... ; 最终第一点的等式(a)左边, z 相关因子 Δ 所有乘积就是:

$$\begin{aligned} & \Delta * \Delta^{-1(C_1^{|G|-|z|})} * \Delta^{-1^2(C_2^{|G|-|z|})} * \dots * \Delta^{-1^{|G|-|z|}(C_{|G|-|z|}^{|G|-|z|})} \\ & = \Delta^{(1 - C_1^{|G|-|z|} + C_2^{|G|-|z|} - \dots + (-1)^{|G|-|z|} C_{|G|-|z|}^{|G|-|z|})} \end{aligned}$$

令 $K = |G| - |z|$, 根据公式(13)可以看出所有的因子互相抵消 $\Delta^0 = 1$; 除了一种情况 $z = G$ 。因而有

$$\prod_{s \subset G} f_s(X_s) = \Delta_{z=G} = P(X_G, X_{G \setminus G} = 0) = P(X_G) = P(X)$$

i. 第一点证明完毕。

证明第二点

为证明第二点，需要使用马尔科夫属性，如果 s 不是一个团，那么一定有两个属于 s 的顶点 a 、 b ，它们之间没有边连接，我们按照如下方式重写 $f_s(X_s)$

$$f_s(X_s = x_s) = \prod_{z \subset s} P(X_z = x_z, X_{G \setminus z} = 0)^{-1^{|s| - |z|}} \quad (14)$$

$$= \prod_{w \subset s \setminus a, b} \left[\frac{P(X_w, X_{G \setminus w} = 0) P(X_{w \cup a, b}, X_{G \setminus w \cup a, b} = 0)}{P(X_w \cup a, X_{G \setminus w \cup a} = 0) P(X_w \cup b, X_{G \setminus w \cup b} = 0)} \right]^{-1^*} \quad (15)$$

公式(15)将 $z \subset s$ 分为4中情况： $z = w, z = w \cup a, z = w \cup b$ 和 $z = w \cup a, b$ ，并显示的写出了这些因子。注意公式(15)中的位置是对的哦。接下来将证明他们互相抵消。因此指数是多少不重要了，这里用 -1^* 表示。

根据贝叶斯规则有：

$$\begin{aligned} & \frac{P(X_w, X_{G \setminus w} = 0)}{P(X_w \cup a, X_{G \setminus w \cup a} = 0)} \\ &= \frac{P(X_a = 0 | X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0) P(X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0)}{P(X_a | X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0) P(X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0)} \end{aligned} \quad (16)$$

$$= \frac{P(X_a = 0 | X_b, X_w, X_{G \setminus w \cup a, b} = 0) P(X_b, X_w, X_{G \setminus w \cup a, b} = 0)}{P(X_a | X_b, X_w, X_{G \setminus w \cup a, b} = 0) P(X_b, X_w, X_{G \setminus w \cup a, b} = 0)} \quad (17)$$

$$= \frac{P(X_w \cup b, X_{G \setminus w \cup b} = 0)}{P(X_w \cup a, b, X_{G \setminus w \cup a, b} = 0)} \quad (18)$$

a) 公式(16)依据概率分解 $P(a, b) = P(a|b) * P(b)$ ；首先仅仅看因子部分

$$P(X_w, X_{G \setminus w} = 0) \setminus = P(X_a = 0, X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0) \setminus = P(X_a = 0 | X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0) P(X_b = 0, X_w, X_{G \setminus w \cup a, b} = 0)$$

b) 同理分母部分一样的分解

c) 公式(16)的左边部分，由于 X_a 和 X_b 在给定向图剩余部分是条件独立的，因此可以将 $X_b = 0$ 替换为 X_b ；分子分母都替换了。

d) 公式(16)的右边部分，分子分母是一样的，可以约掉，实际上是先约掉，然后在同时乘一个相同的因子；得到公式(17)，自然而然概率连乘得到公式(18)

e) 将公式(18)结果带入公式(15)；可知公式(15)恒等于1. 第二点证明完毕。

疑问点

1. 本文的证明，只能说明无向图模型的概率可以分解为 G 上所有团的势函数乘积；并不能说明是所有的最大团的势函数乘积。哪位网友知道，麻烦给我回复，非常感谢！