Statistical Computing
Chap. 4.3: Markov Random Field

LIU, Ran

Department of Statistics,
Beijing Normal University

May 6, 2024

北京師範大學
BEIJING NORMAL UNIVERSITY

Summary

Introduction

马尔可夫随机场模型

马尔可夫随机场是一种图模型，其中节点表示随机变量，边表示变量之间的依赖关系。通过定义节点和边之间的条件概率分布，马尔可夫随机场描述了变量之间的概率关系，使我们能够推断和预测这些变量的状态。

马尔可夫随机场经常应用于图像处理、计算机视觉、自然语言处理等领域，用于建模和分析具有空间相关性的随机变量。

MRF 非常通用，并且可以用于许多晶格类型结构，比如规则矩形、六边形和不规则网格结构。Besag 发布了许多关键论文在 MRF 的空间统计和图像分析上，包括他在 1974 年的开创性论文。

在规则矩形晶格上的应用

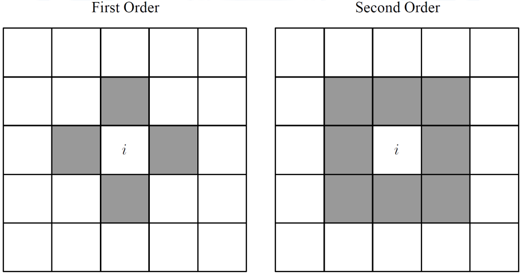
为了简单起见，我们在这里专注于马尔可夫随机场在规则矩形晶格上的应用。

例如，我们可能会在一张地图或图像上叠加一个矩形网格，并且标记晶格中的每个像素或单元格。晶格中第 i 个像素的值用 x_i 表示，对于 $i = 1, \ldots, n$，其中 n 是有限的。

我们将专注于二元随机场，其中 x_i 只能取两个值，0 和 1，对于 $i = 1, \ldots, n$。这些方法也可以简单扩展到 x_i 是连续的或者取超过两个离散值的情况。

邻域定义

让 δ_i 表示一组 x 值，对应于邻近像素 i 的像素值。定义了 δ_i 的像素被称作像素 i 的邻域。像素 x_i 不包含在 δ_i 中。一个合适的邻域定义必须满足条件：如果像素 i 是像素 j 的邻居，则像素 j 也是像素 i 的邻居。



在一个矩形晶格中，一阶邻域是指与目标像素垂直和水平相邻的像素集合。二阶邻域还包括了与目标像素对角线相邻的像素。

局部依赖的马尔可夫随机场

局部依赖的马尔可夫随机场指出，给定剩余像素 X_{-i} 的条件下，像素 X_i 的分布仅依赖于它的邻居像素。因此，对于 $X_{-i}=x_{-i}$，我们有以下条件分布：

$$f(x_i|X_{-i}) = f(x_i|x_{\delta_i}) \tag{1}$$

这里假设每个像素成为 0 或 1 的概率非零，这意味着所谓的正性条件得到满足：即 X 的最小状态空间等于其组成部分的状态空间的笛卡尔积。正性条件确保本节后面考虑的条件分布是良好定义的。

如果某个像素的状态空间为空集，那么条件分布将无法被定义和计算。

举个简单的例子，假设我们有两个随机变量 X 和 Y，它们的取值范围分别为 0, 1 和 a, b。如果我们要计算在观测到 X=0 的情况下，Y=a 的条件概率，我们需要知道 X=0 和 Y=a 同时发生的概率。但是如果 X 的状态空间为空集，表示 X 无法取任何值，那么我们无法计算这个条件概率，因为没有 X=0 和 Y=a 同时发生的可能性。

因此，为了确保条件分布的定义和计算的可行性，我们需要保证相关变量的状态空间都是非空的，至少包含相应的可能取值。这样，条件分布才能在给定观测或已知条件时提供有意义的概率信息。

Hammersley-Clifford 定理表明，条件分布 $f(x)$ 联合指定了 X 的联合分布，除了一个归一化常数。对于我们的离散二元状态空间，这个归一化常数是所有 x 在状态空间上的 $f(x)$ 的和。

这个和通常不可能通过直接计算获得，因为项数是巨大的。即使对于一个不切实际的小的 40×40 像素图像，其二元值的情况下，有 $2^{1600} = 4.4 \times 10^{481}$ 项在求和中。

但利用统计计算算法的优势 (可在除归一化常数情况下采样)，我们能简单对 MRF 进行推断。

Hammersley-Clifford 定理

根据 Hammersley-Clifford 定理，给定一个马尔可夫随机场（Markov Random Field，MRF）的图结构和每个节点的条件分布，我们可以通过势函数的乘积来获得整个随机变量 X 的联合分布。

设 X 为随机变量的集合，对于给定的条件分布 $P(X_i|X_{\delta(i)})$，其中 $X_{\delta(i)}$ 表示与节点 i 相邻的所有节点的集合。我们可以定义一个势函数 $\phi_i(X_i)$ 来表示节点 i 的条件分布，满足 $P(X_i|X_{\delta(i)}) = \phi_i(X_i)/Z_i$，其中 Z_i 是归一化常数。

根据 Hammersley-Clifford 定理，整个随机变量 X 的联合分布可以表示为势函数的乘积：

$$P(X) = (1/Z)\prod_i \phi_i(X_i)$$

其中，$Z = \sum_X \prod_i \phi_i(X_i)$ 是一个归一化常数，用于确保联合分布的总和为 1。

需要注意的是，上述表达中的乘积是在所有节点上进行的，即对于图中的每个节点 i，都存在一个势函数 $\phi_i(X_i)$。这个乘积表示了所有节点的条件分布的乘积。

在上述表达中，我们可以忽略分子中的归一化常数 Z_i，因为它们会在整个联合分布的归一化常数 Z 中相互抵消。因此，在可差一个常数的前提下，我们可以直接使用势函数的乘积来表示整个联合分布。

Ref: Hammersley Clifford Theorem

Hidden Markov Random Fields

假设 y_i 表示第 i 个像素的观测值。因此 X 是一个参数向量，而 y 是数据。在图像分析应用中，y 是降质图像，而 X 是未知的真实图像。

降质图像指的是经过某种损失或噪声处理后的图像，可能包含模糊、噪声、失真等问题。这些降质可能是由于图像获取设备的限制、传输过程中的干扰或其他因素引起的。

在空间统计学中应用到植物或动物物种分布的绘制时，考虑一个像素格点网络，每个像素格点表示地理空间中的一个位置。假设我们对每个像素进行观测，以确定该位置是否存在某种植物或动物物种。$y_i = 0$ 可能表示在第 i 个像素期间未观测到该物种，而 X_i 可能表示该物种在第 i 个像素中的真实（未观测到的）存在或缺失。

在构建这个模型时，有三个假设是基础。首先，我们假定在给定真实像素值的情况下，观测值是相互独立的。所以给定 $X = x$ 时，Y 的联合条件密度为

$$f(y_1, \ldots, y_n | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(y_i | x_i),$$

其中 $f(y_i|x_i)$ 是给定像素 i 的真实值时观测数据 y_i 的密度函数。

模型的隐变量是 $x_1, \ldots, x_n$，分析的目标是估计这些真实值。为此我们采用了吉布斯抽样方法。假设参数的先验分布为 $X \sim f(x)$。因此，在吉布斯抽样中的目标是获得 X 的后验密度样本，

$$f(x|y) \propto f(y|x) f(x). \tag{2}$$

先验密度类别

X 的一个先验密度类别由下式给出

$$f(x) \propto \exp\left(-\sum_{i\sim j}\phi(x_i - x_j)\right),$$

其中 $i \sim j$ 表示所有这样的像素对，即像素 i 是像素 j 的邻居，而 ϕ 是一个关于 0 对称且随着 $|x_i - x_j|$ 增大而增大的函数，其被称为成对差异先验。采用基于成对相互作用的这种先验分布可以简化计算。

吉布斯抽样器

吉布斯抽样器需要导出单变量条件分布的形式。因此，在迭代 t 时的吉布斯更新为

$$X_i^{(t+1)} | X_{-i}^{(t)} \sim f\left(x_i | x_{-i}^{(t)}, y\right). \tag{3}$$

简单的例子

设 $x_i = 1$ 表示物种确实存在于像素 i 中。我们考虑由数据密度推导出的简单似然函数：

$$f(y|x) \propto \exp\left(\alpha \sum_{i=1}^{n} \mathbb{I}\{y_i = x_i\}\right), \tag{4}$$

这里 $x_i \in \{0, 1\}$。参数 α 可以是用户选择的常数，或者通过为其选择一个先验来估计。我们在这里采取前者的方法，设置 $\alpha = 1$。

我们假设 X 的成对差异先验密度由下式给出：

$$f(x) \propto \exp\left(\beta \sum_{i \sim j} \mathbb{I}\{x_i \neq x_j\}\right),$$

这里 $x \in S = \{0, 1\}^{46 \times 54}$。我们考虑第一阶邻域，所以 $i \sim j$ 在上式中表示对水平和垂直邻近的像素进行求和，对于 $i = 1, \ldots, n$。方程引入了超参数 β，可以为其分配一个超先验或视为常数。通常 β 是被视为一个常数，由用户定义。

条件分布和吉布斯抽样更新

在此，我们设定 $\beta = 0.8$ 来促进类似颜色像素的聚类。为了确定所选择的 α 和 β 值的效果，建议进行灵敏度分析。

X_i 给定 X_{-i}, y 的单变量条件分布是伯努利分布。因此，在吉布斯抽样的 $(t+1)$ 次循环中，第 i 个像素被设为 1 的概率为

$$P(X_i^{(t+1)} = 1 | X_{-i}^{(t)}, y) = \Bigg(1 + \exp\Bigg\{\alpha \left(\mathbb{I}\{y_i = 0\} - \mathbb{I}\{y_i = 1\}\right) + \beta \sum_{i \sim j} \left(\mathbb{I}\{x_j^{(t)} = 0\} - \mathbb{I}\{x_j^{(t)} = 1\}\right)\Bigg\}\Bigg)^{-1},$$

对于 $i = 1, \ldots, n$。

这里使用的模型是基本的，忽略了在分析此类空间数据时可能出现的许多重要问题。例如，当通过合并空间参考数据来创建像素时，如果观察到物种某些部分存在于像素 i 中，某些在其他像素中，则不清楚如何对像素 i 的观察到的响应进行编码。

当然在实际过程中，一个像素可能包含的是一个概率 p_i(某事件发生的概率)

$$p(y_i = 1 \mid x_i = 1) = p_i$$

结合第 i 个像素的协变量 w_i，我们可建立

$$\log(\frac{p_i}{1-p_i}) = w_i^T \beta + \nu_i$$

其中 β 为协变量的系数变量，ν_i 为空间相关的随机相应。(空间流行病学)