

<https://www.cnblogs.com/simayuhe/p/5444091.html>

马尔科夫随机场(MRF)及其在图像降噪中的matlab实现

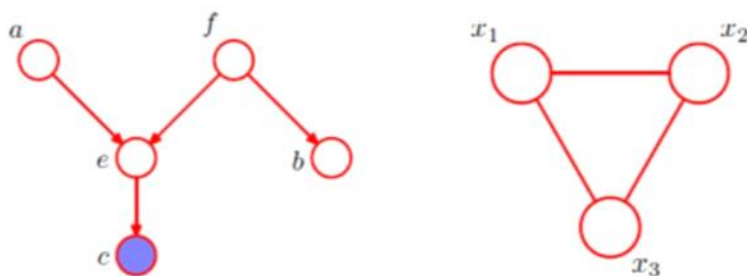
(Markov Random Field) 马尔科夫随机场，本质上是一种概率无向图模型

下面从概率图模型说起，主要参考PR&ML 第八章 Graphical Model (图模型)

定义：A graph comprises nodes (also called vertices) connected by links (also known as edges or arcs).

In a probabilistic graphical model each node represents a random variable or group of random variables ,and the links express probabilistic relationships between these variables.

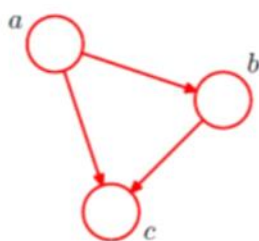
分类：概率图模型大体分为两种，第一种是有向图，又叫贝叶斯网络，链接节点之间的边是有方向的，用来描述变量之间的因果关系（左侧）；另一种是无向图，又叫马尔科夫随机场，边是没有方向的，用来描述变量之间的软约束（右侧）。



1. Bayesian network

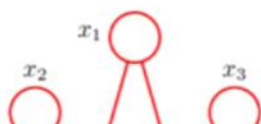
用概率图表示联合概率分布密度

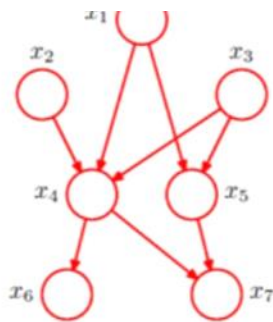
$$p(a, b, c) = p(c | a, b)p(b | a)p(a)$$



起点对应条件变量，如 $p(c|a,b)$ 就会有二个起点一个终点。略微复杂一点如下图：

$$p(x_1)p(x_2)p(x_3)p(x_4 | x_1, x_2, x_3)p(x_5 | x_1, x_3)p(x_6 | x_4)p(x_7 | x_4, x_5)$$





由于变量之间如果采用全连接会使得参数个数随着变量个数成幂次增加。所以书中讨论了几种在设计贝叶斯网络时减少模型参数的方法，第一种是独立性假设，即根据实际问题对某些变量提出独立性假设以减少链接边数；第二种是采用链式连接，即每个节点只有一个子节点一个父节点；第三种是权值共享，类似于cnn中底层的做法。

2 condition independence 条件独立

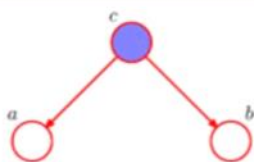
定义式：

$$\begin{aligned} p(a, b | c) &= p(a | b, c)p(b | c) \\ &= p(a | c)p(b | c) \end{aligned}$$

下面来讨论如何通过图结构直接判断变量之间的条件独立性

三个例子，一段总结

例一：结点c被称为关于（从a经过c到达b）这个路径的"尾到尾"（tail-to-tail）连接，因为结点与两个箭头的尾部相连

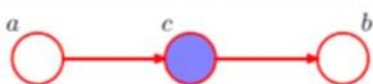


注意图中c是可观测量

$$\begin{aligned} p(a, b | c) &= \frac{p(a, b, c)}{p(c)} \\ &= p(a | c)p(b | c) \end{aligned}$$

记作 $a \perp\!\!\!\perp b | c$

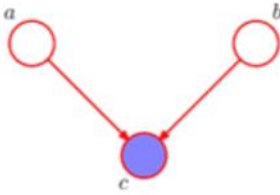
例二：结点c被称为关于从结点a到结点b的路径"头到尾"（head-to-tail）。



$$p(a, b | c) = \frac{p(a, b, c)}{p(c)}$$

$$\begin{aligned}
 &= \frac{p(a)p(c|a)p(b|c)}{p(c)} \\
 &= p(a|c)p(b|c)
 \end{aligned}$$

例三：结点c关于从a到b的路径是"头到头"（head-to-head），因为它连接了两个箭头的头。



$$\begin{aligned}
 p(a, b | c) &= \frac{p(a, b, c)}{p(c)} \\
 &= \frac{p(a|c)p(b|c)p(c|a, b)}{p(c)}
 \end{aligned}$$

注意这里当c为观测条件时不能得到 $p(a)p(b)$ 的形式，所以 $a \not\perp b | c$

一个总结：In summary, a tail-to-tail node or a head-to-tail node leaves a path unblocked unless it is observed in which case it blocks the path. By contrast, a head-to-head node blocks a path if it is unobserved, but once the node, and /or at least one of its descendants, is observed the path becomes unblocked. 参考译文（总而言之，TT结点或者HT结点形成的路径无堵，除非该结点被观察到（这会堵住路径）。相反，HH结点在没被观察到时会堵住路径，不过一旦该结点或其至少一个子结点被观察到的话，路径又变成无堵的了～）

一个小概念 D-separation：这个D-separation的目的是通过有向无环图寻找一个特定的

条件依赖表述 $a \perp b | c$ 。针对节点（变量）集合，如果所有从A经过C到达B的路径都被阻隔，那么我们说C把A从B中d-separation. 即 $A \perp B | C$ 。

3. Markov Random Fields

一个马尔科夫随机场（Markov random field），也被称为马尔科夫网络（Markov network）或者无向图模型（undirected graphical model）

（Kindermann and Snell, 1980），包含一组结点，每个结点都对应着一个变量或一组变量。链接是无向的，即不含有箭头。

这里同样来讨论他的独立性，集合A到集合B的所有路径都通过集合C中的一个或多个节点，那么所有这样的路径都被阻隔，因此 $A \perp B | C$ 。或者说，除掉集合C之后A，B就没有连接路径了。

注意，只要有一条路径没有通过C，那么这种性质就不存在。

定义团块 (clique) : 对这个团块的所有子集的节点都是全连接的。

$$p(x) = \frac{1}{Z} \prod_C \psi_C(x_C)$$

那么整个图的联合分布由各个团块组成,

这里为每一个团块定义了一个势函数, 注意如果这个无向图不是由有向图转化而来, 那么这里的势函数可以不满足概率性质。但是, 由于 $p(x)$ 要满足概率性质, 所以

势函数还是要大于零的。 Z 是归一化因子。通常取 $\psi_C(x_C) = \exp\{-E(x_C)\}$ 这个分布叫做 Boltzmann distribution (有时间要查查他的性质)。而这里最大的优势就是势函数可以通过选取而具有一定的实际意义, 相当于一种相似度的度量, 所以为算法的设计增加了灵活性。

例子: 图像去噪



这个例子可以有很多更好的方法实现, 这里只是为了说明势函数的用法而已。

给一幅二值图像加入噪声, 加入噪声之后的图像是我们的观测值 Y , 而实际图像是隐变量 X , 现在我们要通过 Y 推断 X ,

由于噪声等级比较小, 因此我们知道 x_i 和 y_i 之间有着强烈的相关性。我们还知道图像中相邻像素 x_i 和 x_j 的相关性很强。这种先验知识可以使用马尔科夫随机场模型进行描述, 它的无向图如下:

势函数定义为:

$$E(x, y) = h \sum_i x_i - \beta \sum_{\{i, j\}} x_i x_j - \eta \sum_i x_i y_i$$

这里只有两种团块, 第一种是 x_i 和 y_i , 第二种是 x_i 和相邻的 x_j 。我们分别给以不同系数

用来调节其在势函数中的权重, 由于最后的概率计算要取

$$p(x, y) = \frac{1}{Z} \exp\{-E(x, y)\}$$

最大, 所以这里等价于取 $E(x, y)$ 最小。

具体代码如下:

```
1 %实现PRML P387 例子: 使用mrf降噪
2 %输入一个二值图像并加入噪声
3 close all;
4 clear all;
```



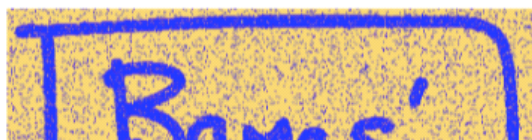
```

4 I=imread('Penguins.jpg');
5 I1=im2bw(I);
6 subplot(2,2,1)
7 imshow(I1);
8 title('原图像');
9
10 J = imnoise(I,'salt & pepper', 0.2);
11 J1=im2bw(J);
12 subplot(2,2,2)
13 imshow(J1);
14 title('噪声图')
15
16 Y=ones(size(J1));
17 Y(J1==0)=-1;
18 [m,n]=size(Y);
19 X=Y;h=0;beta=3.5;eta=.1;
20 while 1
21     tot=0
22     for i=2:1:m-1
23         for j=2:1:n-1
24             temp=X(i,j);
25             X(i,j)=-1;%根据定义计算势函数
26             E1=h*X(i,j)-beta*X(i,j)*(X(i-1,j)+X(i+1,j)+X(i,j-1)+X(i,j+1))-eta*X(i,j)*Y(i,j);
27             X(i,j)=1;
28             E2=h*X(i,j)-beta*X(i,j)*(X(i-1,j)+X(i+1,j)+X(i,j-1)+X(i,j+1))-eta*X(i,j)*Y(i,j);
29             if E1<E2
30                 X(i,j)=-1;
31             else
32                 X(i,j)=1;
33             end
34             if temp~=X(i,j)
35                 tot=tot+1;
36             end
37         end
38     end
39     if tot<1
40         break;
41     end
42 end
43
44 J2=X;
45 J2(X==-1)=0;
46 subplot(2,2,3)
47 imshow(J2);
48 title('mrf降噪结果')

```

二值图像去噪实际做的。
有一张有噪声的图像能反恢复。

我们可以使用二值图像中图像去噪的例子来说明无向图的应用 (Besag, 1974; Geman and Geman, 1984; Besag, 1986)。虽然这是一个非常简单的例子，但是它可以代表许多更复杂的应用。我们令观测的噪声图像通过一个二值像素值 $y_i \in \{-1, +1\}$ 组成的数组来描述，其中下标 $i = 1, \dots, D$ 覆盖了所有的像素。我们假设图像通过取一张未知的无噪声图像，这幅图像由二值像素值 $x_i \in \{-1, +1\}$ 描述，然后以一个较小的概率随机翻转像素值的符号的方式获得。图8.30给出了一个二值图像以及一副噪声图像，其中噪声图像中像素值发生符号翻转的概率为10%。



我们可以使用二值图像中图像去噪的例子来说明无向图的应用 (Besag, 1974; Geman and Geman, 1984; Besag, 1986)。虽然这是一个非常简单的例子, 但是它可以代表许多更复杂的应用。我们令观测的噪声图像通过一个二值像素值 $y_i \in \{-1, +1\}$ 组成的数组来描述, 其中下标 $i = 1, \dots, D$ 覆盖了所有的像素。我们假设图像通过取一张未知的无噪声图像, 这幅图像由二值像素值 $x_i \in \{-1, +1\}$ 描述, 然后以一个较小的概率随机翻转像素值的符号的方式获得。图8.30给出了一个二值图像以及一副噪声图像, 其中噪声图像中像素值发生符号翻转的概率为10%。



图 8.30 使用马尔科夫随机场进行图像去噪的例子。上面一行中, 左侧是原始的二值图像, 右侧是随机改变 10% 的像素后得到的带有噪声的图像。下面一行中, 左图是使用迭代条件模型(ICM)恢复的图像, 右图是使用最大割算法得到的图像。ICM产生的图像中, 96%的像素与原始图像相符, 而最大割算法产生的图像中, 这个比例为99%。

给定带有噪声的图像, 我们的目标是恢复原始的无噪声的图像。

由于噪声等级比较小, 因此我们知道 x_i 和 y_i 之间有着强烈的相关性。我们还知道图像中相邻像素 x_i, x_j 的相关性很强。这种先验知识可以使用马尔科夫随机场模型进行描述, 它的无向图如图8.31所示。

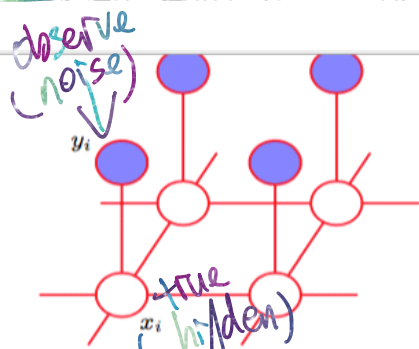


图 8.31 一个无向图模型, 表示图像去噪的马尔科夫随机场, 其中 x_i 是一个二值变量, 表示像素 i 在一个未知的无噪声的图像中的状态, y_i 表示在观测到的噪声图像中, 像素 i 的对应值。

这个图中有两种类型的团块, 每一种团块包含两个变量。形如 $\{x_i, y_i\}$ 的团块有一个关联的能量函数, 表达了这些变量之间的相关性。对于这些团块, 我们选择一个非常简单的能量函数 $-\eta x_i y_i$, 其中 η 是一个正的常数。这个能量函数具有当 x_i, y_i 符号相同时, 能量函数会给出一个较低的能量 (即较高的概率), 而当 x_i, y_i 符号相反时, 能量函数会给出一个较高的能量的效果。

剩余的团块由变量 $\{x_i, x_j\}$ 组成, 其中 i, j 是相邻像素的下标。与之前一样, 我们希望当两个像素符号相同时

符号相反时, 能量函数会给出一个较高的能量的效果。

剩余的团块由变量 $\{x_i, x_j\}$ 组成, 其中 i, j 是相邻像素的下标。与之前一样, 我们希望当两个像素符号相同时能量较低, 当两个像素符号相反时能量较高, 因此我们选择能量函数 $-\beta x_i x_j$, 其中 β 是一个正的常数。

由于势函数是最大团块上的一个任意的非负的函数, 因此我们可以将势函数与团块的子集上的任意非负函数相乘, 或等价的, 我们可以加上对应的能量。在这个例子中, 这使得我们可以为无噪声图像的每个像素 i 加上一个额外的项 $h x_i$ 。这样的项具有将模型进行偏置, 使得模型倾向于选择一个特定的符号, 而不选择另一个符号的效果。

于是, 模型的完整的能量函数的形式为

$$E(x, y) = h \sum_i x_i - \beta \sum_{\{i, j\}} x_i x_j - \eta \sum_i x_i x_j \quad (8.42)$$

它定义了 x, y 上的一个联合概率分布, 形式为

$$p(x, y) = \frac{1}{Z} \exp\{-E(x, y)\} \quad (8.43)$$

我们现在固定噪声图像的像素给出的观测值 y 的元素, 这个噪声图像隐式地定义了一个无噪声图像上的条件概率分布 $p(x|y)$ 。这是Ising模型的一个例子, 这个模型在统计物理学中被广泛研究。为了恢复图像, 我们希望找到一个具有较高概率 (理想情况下具有最高概率) 的图像 x 。为了完成这件事, 我们要使用一个简单的迭代方法, 叫做迭代条件峰值 (iterated conditional model), 或称为ICM (Kittler and Foglein, 1984)。这种方法仅仅是坐标间的梯度上升方法的一个应用。这种方法的思想是, 首先初始化变量 $\{x_i\}$, 这个过程中我们只是简单地令 $x_i = y_i$ 对于所有 i 都成立。然后, 我们每次取一个 x_i 结点, 计算两个可能状态 $x_i = +1$ 和 $x_i = -1$ 的总能量, 保持其他所有结点变量固定, 将 x_i 设置为能量较低的状态。如果 x_i 不变, 则概率不变, 否则概率就会增大。由于只有一个变量发生改变, 因此这是一个可以高效进行的简单局部计算。然后, 我们对其他的结点重复更新过程, 以此类推, 知道满足某个合适的停止条件。结点可以用一种系统的方式更新, 例如重复地依次扫描图像, 或随机地选择结点。

如果我们有一个更新的顺序, 使得每个像素都至少被访问一次, 且没有变量发生改变, 那么根据定义, 算法会收敛于概率的一个局部最大值。然而, 这未必对应于全局最大值。

对于这个简单的例子来说, 我们将参数固定为 $\beta = 1.0, \eta = 2.1, h = 0$ 。注意, 令 $h = 0$ 意味着两个状态 x_i 的先验概率是相等的。首先, 我们使用噪声图像进行初始化, 然后运行ICM直到收敛, 得到了图8.30左下角的去噪图像。注意, 如果我们令 $\beta = 0$, 从而去除了相邻像素点之间的联系, 那么整体概率最大的解为 $x_i = y_i$ (对于所有的 i), 这对应于观测到的噪声图像。

Mean field inference for binary images The MNIST dataset consists of 60,000 images of handwritten digits, curated by Yann LeCun, Corinna Cortes, and Chris Burges. You can find it [here](#), together with a collection of statistics on recognition, etc. We will use the first 500 of the training set images.

- Obtain the MNIST training set, and binarize the first 500 images by mapping any value below .5 to -1 and any value above to 1. For each image, create a noisy version by randomly flipping 2% of the bits.
- Now denoise each image using a Boltzmann machine model and mean field inference. Use $\theta_{ij} = 0.2$ for the H_i, H_j terms and $\theta_{ij} = 2$ for the H_i, X_j terms. **To hand in:** Report the fraction of all pixels that are correct in the 500 images.
- To hand in:** Prepare a figure showing the original image, the noisy image, and the reconstruction for
 - your most accurate reconstruction
 - your least accurate reconstruction
- Assume that θ_{ij} for the H_i, H_j terms takes a constant value c . We will investigate the effect of different values of c on the performance of the denoising algorithm. Think of your algorithm as a device that accepts an image, and for each pixel, predicts 1 or -1. You can evaluate this in the same way we evaluate a binary classifier, because you know the right value of each pixel. A **receiver operating curve** is a curve plotting the true positive rate against the false positive rate for a predictor, for different values of some useful parameter. We will use c as our parameter. **To hand in:** Using $c = (-1, 0, 0.2, 1, 2)$ plot a receiver operating curve for your denoising algorithm.

- ① 读取前500个数据
- ② Binarize 图片, 使之二值图像
- ③ Add noise: 随机反转%的像素, 作为noise
- ④ Denoise: 用 Boltzmann machine model 和 Mean field Inference 去噪.