第7章：概率图模型和深度生成模型

中国科学技术大学
电子工程与信息科学系

主讲教师：李厚强 (lihq@ustc.edu.cn)
周文罡 (zhwg@ustc.edu.cn)
李　礼 (lil1@ustc.edu.cn)
胡　洋 (eeyhu@ustc.edu.cn)

概率图模型（Probabilistic Graphical Models）

- 概率图模型概述
- 概率有向图模型（贝叶斯网络）
  - 因子分解
  - 条件独立性
- 概率无向图模型（马尔可夫随机场）
  - 因子分解
  - 条件独立性
- 信念传播算法
- 应用举例

概率建模

- 图像处理和分析：观测数据 $\xrightarrow{\text{推测}}$ 未知数据
- 不确定性
  - 观测不确定性
  - 预测不确定性
- 概率模型
  - 链式法则：$P(x_1,\cdots,x_N)=P(x_1)\prod_{i=2}^{N}P(x_i|x_1,\cdots,x_{i-1})$
  - 条件独立性假设
- 概率图模型
  - 概率模型
  - 引入图作为表示工具

概率图模型

□ 模型定义

- 图（graph）是由结点（node）和边（edge）组成的集合
  - ✓ 结点：v，结点的集合：V
  - ✓ 边：e，边的集合记：E
  - ✓ 图：$G = (V, E)$
- 概率图模型
  - ✓ 结点v，随机变量Y_v
  - ✓ 边e：随机变量之间的概率依赖关系
  - ✓ 用图$G = (V, E)$表示联合概率分布$P(Y)$

□ 概率图模型的优点

- 使概率模型的结构可视化，启发新模型设计
- 便于分析模型的各种性质，如条件独立性
- 将复杂计算（如推断、学习问题）表示成图操作

概率图模型的基本问题

- 表示问题
  - 概率有向图模型（贝叶斯网络）
  - 概率无向图模型（马尔可夫随机场）
- 学习问题
  - 参数学习
  - 图结构学习
- 推断问题

概率图模型（Probabilistic Graphical Models）

- □ 概率图模型概述
- □ 概率有向图模型（贝叶斯网络）
  - ■ 因子分解
  - ■ 条件独立性
- □ 概率无向图模型（马尔可夫随机场）
  - ■ 因子分解
  - ■ 条件独立性
- □ 信念传播算法
- □ 应用举例

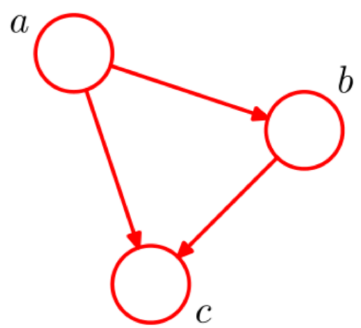
概率有向图模型（贝叶斯网络）

- 有向图表示随机变量之间的因果关系
- 考虑随机变量a, b, c的联合分布$p(a, b, c)$

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

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

- 用有向图模型表示联合分布



- 每个随机变量用一个结点表示
- 根据每个条件分布，在结点间引入有向边，表示随机变量之间的概率依赖关系
- 不同的分解方式，对应不同的图表示

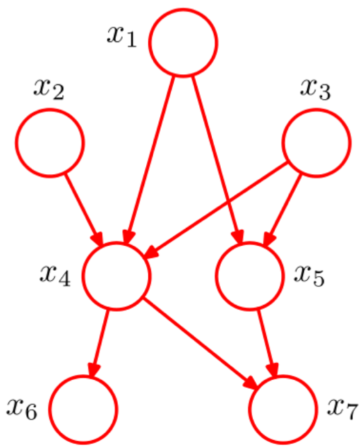
概率有向图模型的因子分解

- 概率有向图表示的联合概率分布可表示为图中每个结点在其父节点给定时的条件分布的乘积

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i|\mathrm{pa}_i)$$

其中pa_i为结点 x_i的父结点集合

要求图中没有有向环，即为有向无环图



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

缺失的边有意义

概率有向图模型的因子分解

- M个结点的Markov链

$$p(\mathbf{x}_1, \cdots, \mathbf{x}_M) = p(\mathbf{x}_1) \prod_{i=2}^{M} p(\mathbf{x}_i | \mathbf{x}_{i-1})$$

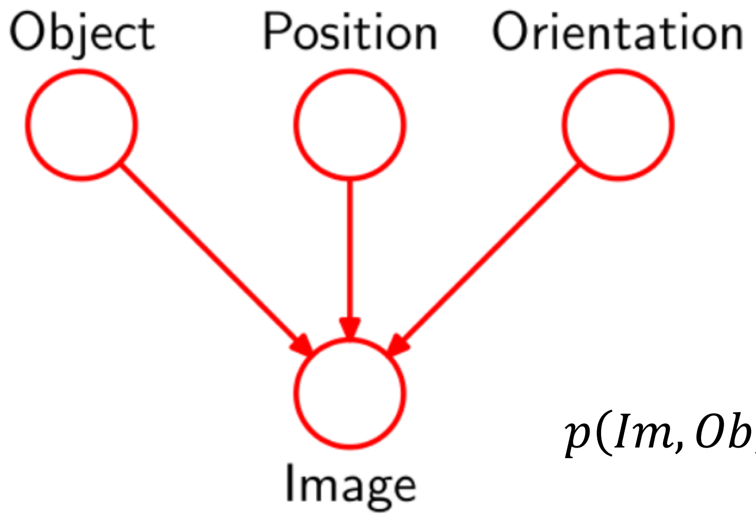
  - $p(\mathbf{x}_1)$有 $K-1$个参数
  - $p(\mathbf{x}_i | \mathbf{x}_{i-1})$有 $K(K-1)$个参数
  - 总共有$K-1+(M-1)(K-1)K$个参数

- 隐马尔可夫模型的核心结构

生成式模型

- 通常以能观察到的变量为叶结点，其他结点为隐变量。通过引入隐变量，使观察变量复杂的分布可以通过更简单的条件分布构造而成。

- 简单的图像生成模型



  - 物体、位置、方向满足相互独立的先验分布
  - 图像的分布以物体、位置、方向为条件

$$p(Im, Ob, Po, Or) = p(Im|Ob, Po, Or)p(Ob)p(Po)p(Or)$$

条件独立性

- 条件独立性质
  - 给定随机变量c的条件下，随机变量a和b独立

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

或者

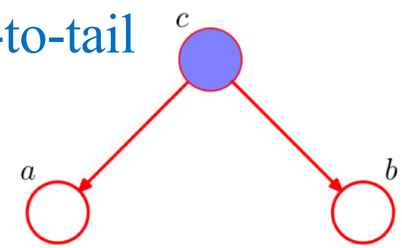
$$\begin{aligned} p(a,b|c) &= p(a|b,c)p(b|c) \\ &= p(a|c)p(b|c) \end{aligned}$$

  - 记为：$a \perp b|c$
- 条件独立性是概率模型的重要性质，既使模型结构更简单，也使模型的推断和学习问题更简单
- 概率图模型的一大优点是可从图中直接得到变量间的条件独立关系

条件独立性

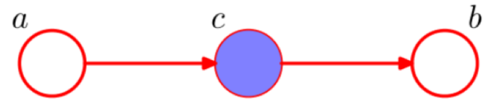
- □ 变量a, b, c的关系

tail-to-tail



$a \perp b|c$

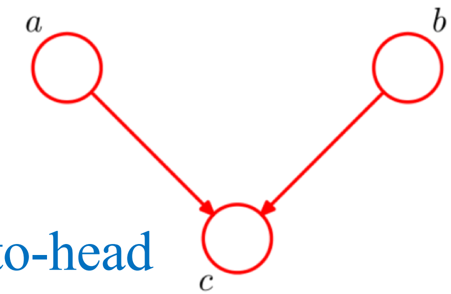
head-to-tail



$a \perp b|c$

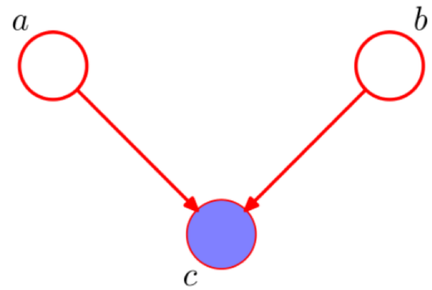
$$p(a,b|c) = \frac{p(c)p(a|c)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

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



$a \perp b|\emptyset$

head-to-head



$a \not\perp b|c$

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

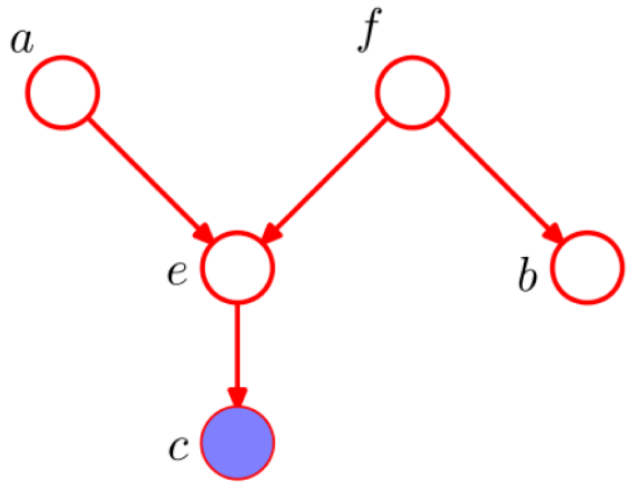
$$p(a,b|c) = \frac{p(a)p(b)p(c|a,b)}{p(c)}$$

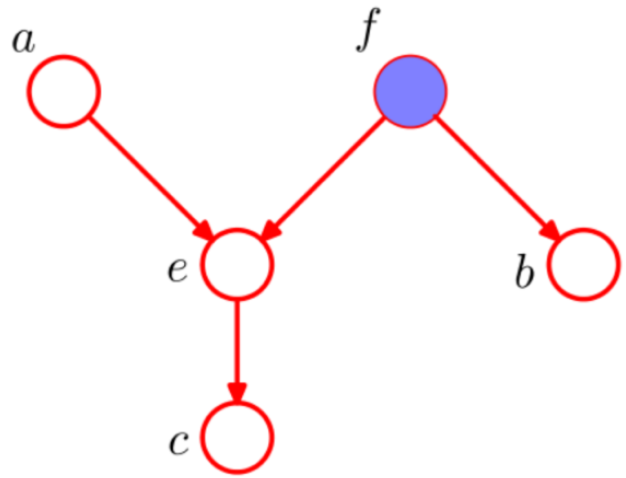
条件独立性

- D-separation性质
  - 有向图中的三个结点集合A, B, C
  - 考虑A中结点到B中结点的所有可能路径
  - 一条路径如果含有一个结点满足下列条件之一，则该路径被阻隔
    - ✓ 路径经过结点时为head-to-tail或tail-to-tail形式，并且该结点在集合C中
    - ✓ 路径经过结点时为head-to-head形式，并且该结点及其子孙结点都不在集合C中
  - 如果A, B间都所有路径都被阻隔，则称A, B被C集合D分离，它们对应的随机变量满足$A \perp B|C$

条件独立性

- 例子





a, b没有被f阻隔

a, b没有被e阻隔

$a \perp b | c$不成立

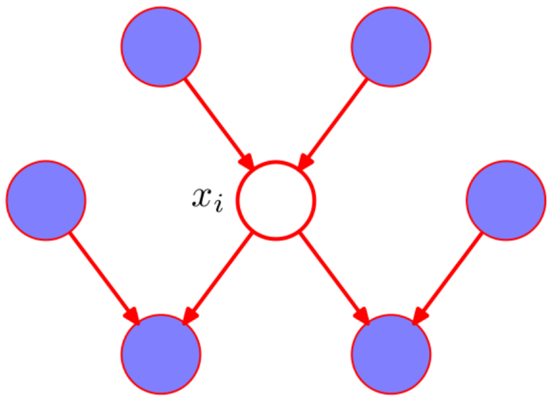
a, b被f阻隔

a, b被e阻隔

$a \perp b | f$成立

有向图模型的马尔可夫毯

- 马尔可夫毯：一个结点的马尔可夫毯是使该结点与图中所有其他结点独立所需要观察到的最小结点集合
- 有向图中，结点的马尔可夫毯包含其父结点、子结点及所有co-parent结点



$$p\left(x_i \middle| x_{\{j\neq i\}}\right) = \frac{p(x_1,\cdots,x_M)}{\int p(x_1,\cdots,x_M)\mathrm{d}x_i}$$

$$= \frac{\prod_k p(x_k|\mathrm{pa}_k)}{\int \prod_k p(x_k|\mathrm{pa}_k)\,\mathrm{d}x_i}$$

不含有x_i的因子上下消掉

$$= p\left(x_i \middle| x_{\{j\in MB_i\}}\right)$$

也可由D-separation性质得到

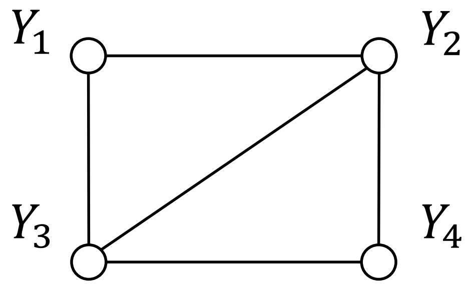
概率图模型（Probabilistic Graphical Models）

- 概率图模型概述
- 概率有向图模型（贝叶斯网络）
  - 因子分解
  - 条件独立性
- 概率无向图模型（马尔可夫随机场）
  - 因子分解
  - 条件独立性
- 信念传播算法
- 应用举例

概率无向图模型的因子分解

- 团与最大团
  - 无向图G中任何两个结点均有边连接的结点子集称为团（clique）
  - 若C是无向图G的一个团，并且不能再加进任何一个G的结点使其成为一个更大的团，则称此C为最大团（maximal clique）

  - 2个结点的团：
    - ✓ $\{Y_1, Y_2\}$，$\{Y_2, Y_3\}$，$\{Y_3, Y_4\}$，$\{Y_4, Y_2\}$，$\{Y_1, Y_3\}$
  - 3个结点的团
    - ✓ $\{Y_1, Y_2, Y_3\}$，$\{Y_2, Y_3, Y_4\}$
  - 最大团
    - ✓ $\{Y_1, Y_2, Y_3\}$，$\{Y_2, Y_3, Y_4\}$



概率无向图模型的因子分解

- 概率无向图模型的因子分解（Factorization）
  - 将概率无向图模型的联合概率分布表示成最大团上的随机变量的函数的乘积形式
- Hammersley-Clifford定理

如果一个联合概率分布$P(Y)$满足无向图G包含的条件独立性，当且仅当 $P(Y)$ 可以表示为一系列定义在最大团上的非负函数的乘积的形式，即

$$P(Y) = \frac{1}{Z}\prod_{C}\psi_C(Y_C)$$

其中C为G上的最大团， Y_C表示C对应的随机变量，$\psi_C(Y_C) \geq 0$是定义在C上的势函数， Z为规范化因子，用来将乘积归一化为概率形式：

$$Z = \sum_{Y}\prod_{C}\psi_C(Y_C)$$

概率无向图模型的因子分解

- 势函数

$$\psi_C(Y_C) = \exp\{-E(Y_C)\}$$

  - ✓ $E(Y_C)$被称为能量函数（energy function）
  - ✓ 总能量可以通过将每个最大团的能量相加的方法得到
  - ✓ 指数表示被称为玻尔兹曼分布（Boltzmann distribution）

$$P(Y) = \frac{1}{Z}\prod_C \exp\{-E(Y_C)\} = \frac{1}{Z}\exp\left\{-\sum_C E(Y_C)\right\}$$

- 规范化因子（划分函数）
  - ✓ 划分函数的计算复杂度是指数级的
  - ✓ 对于参数学习来说，划分函数是必要的；对于局部条件概率分布的计算，划分函数不是必要的

条件独立性

- 成对马尔可夫性
  - 设u和v是无向图G中任意两个没有边连接的结点，其他所有结点为O，对应的随机变量分别为Y_u，Y_v和Y_O
  - 给定随机变量组Y_O的条件下，随机变量Y_u和Y_v是条件独立的

$$P(Y_u, Y_v | Y_O) = P(Y_u | Y_O) P(Y_v | Y_O)$$

- 局部马尔可夫性
  - 设v是无向图G中任意结点，W是与v有边相连的所有结点，O是其他所有结点，对应的随机变量分别为Y_v，Y_W和Y_O
  - 给定随机变量组Y_W的条件下，Y_v与Y_O是条件独立的

$$P(Y_v, Y_O | Y_W) = P(Y_v | Y_W) P(Y_O | Y_W)$$

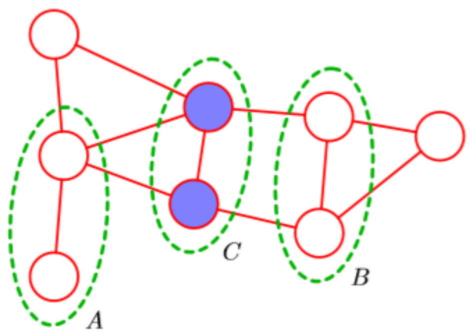
或 $$P(Y_v | Y_W, Y_O) = P(Y_v | Y_W)$$

  - 结点集合W是结点v马尔可夫毯

条件独立性

- 全局马尔可夫性
  - A，B，C是在无向图G中的结点集合，对应的随机变量分别为Y_A，Y_B，Y_C
  - 结点集合A，B被结点集合C分开：
    - ✓ 考虑连接集合A的结点和集合B的结点的所有可能路径
    - ✓ 如果所有这些路径都通过了集合C中的一个或多个结点，那么所有这样的路径都被阻隔
  - 给定随机变量组Y_C的条件下，Y_A和Y_B是条件独立的

$$P(Y_A, Y_B | Y_C) = P(Y_A | Y_C) P(Y_B | Y_C)$$

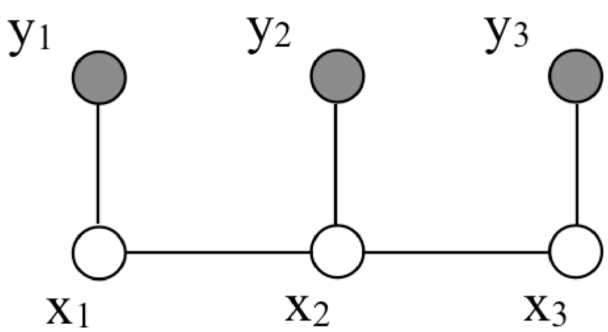


概率图模型（Probabilistic Graphical Models）

- 概率图模型概述
- 概率有向图模型（贝叶斯网络）
  - 因子分解
  - 条件独立性
- 概率无向图模型（马尔可夫随机场）
  - 因子分解
  - 条件独立性
- 信念传播算法
- 应用举例

信念传播算法（Belief Propagation）

- 3结点马尔可夫链



$$p(x_1, x_2, x_3 \mid y_1, y_2, y_3) = \frac{1}{Z(\mathbf{y})}\phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\psi_1(y_1, x_1)\psi_2(y_2, x_2)\psi_3(y_3, x_3)$$

信念传播算法（BP）

□ 边缘概率计算

$$
\begin{aligned}
p(x_1 \mid \mathbf{y}) &= \sum_{x_2}\sum_{x_3} p(x_1, x_2, x_3 \mid \mathbf{y}) \\
&= \frac{1}{Z(\mathbf{y})} \sum_{x_2}\sum_{x_3} \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\psi_1(y_1, x_1)\psi_2(y_2, x_2)\psi_3(y_3, x_3) \\
&= \frac{1}{Z(\mathbf{y})} \psi_1(y_1, x_1) \sum_{x_2} \phi_{12}(x_1, x_2)\psi_2(y_2, x_2) \sum_{x_3} \phi_{23}(x_2, x_3)\psi_3(y_3, x_3)
\end{aligned}
$$

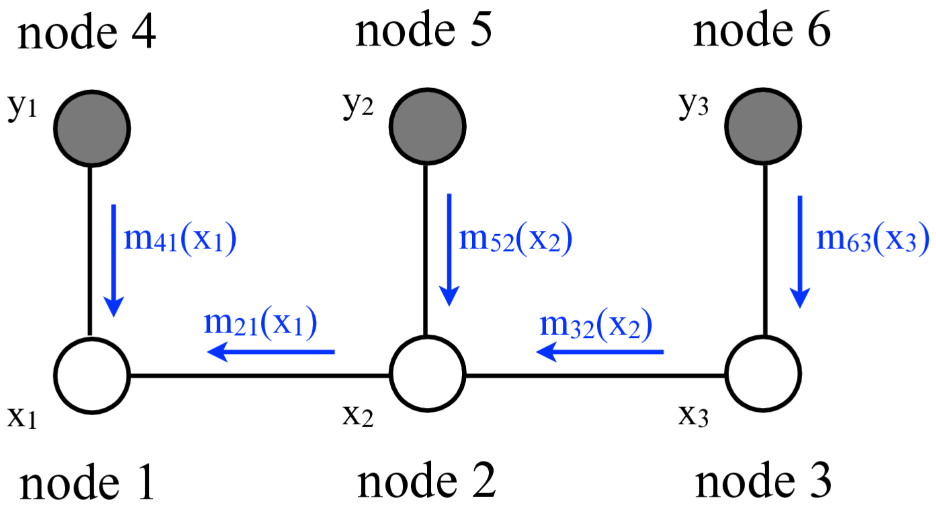
$|x|^3$项加和

$2|x|^2$项加和

对N结点马尔可夫链，由$|x|^N$项加和减少为 $(N-1)|x|^2$项

信念传播算法（BP）

□ 消息传递（Message Passing）



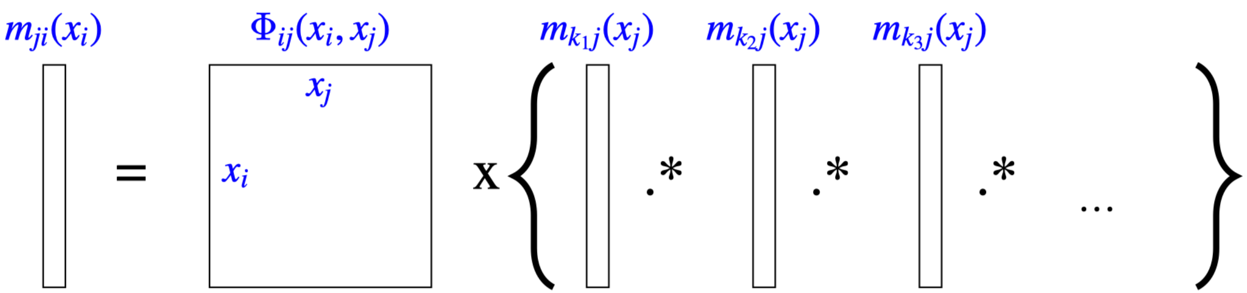
$$
\begin{aligned}
m_{63}(x_3) &= \psi_3(y_3, x_3) \\
m_{52}(x_2) &= \psi_2(y_2, x_2) \\
m_{41}(x_1) &= \psi_2(y_1, x_1) \\
m_{32}(x_2) &= \sum_{x_3} \phi_{23}(x_2, x_3) m_{63}(x_3) \\
m_{21}(x_1) &= \sum_{x_2} \phi_{12}(x_1, x_2) m_{52}(x_2) m_{32}(x_2)
\end{aligned}
$$

$$
p(x_1 \mid \mathbf{y}) = \frac{1}{Z(\mathbf{y})} \psi_1(y_1, x_1) \sum_{x_2} \phi_{12}(x_1, x_2) \psi_2(y_2, x_2) \sum_{x_3} \phi_{23}(x_2, x_3) \psi_3(y_3, x_3)
$$

信念传播算法（BP）

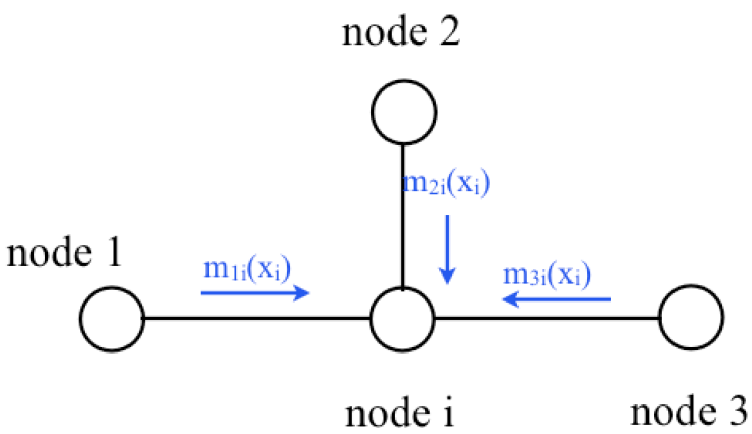
- 结点 j 向结点 i 传递的消息
  - 将除结点 i 外，所有其他结点传递给结点 j 的消息相乘
  - 再与结点 i 和结点 j 之间的势函数相乘
  - 对结点 j 对应的随机变量的所有取值求和

$$m_{ji}(x_i) = \sum_{x_j} \psi_{ij}(x_i, x_j) \prod_{k \in \eta(j) \backslash i} m_{kj}(x_j)$$



信念传播算法（BP）

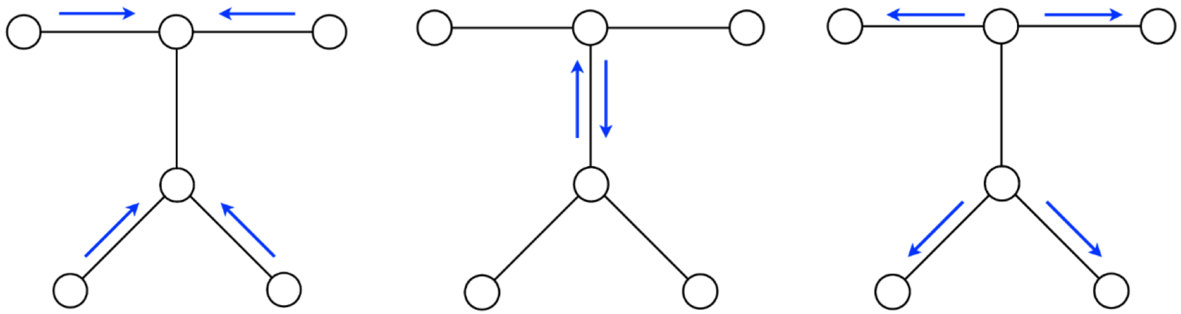
□ 边缘概率计算



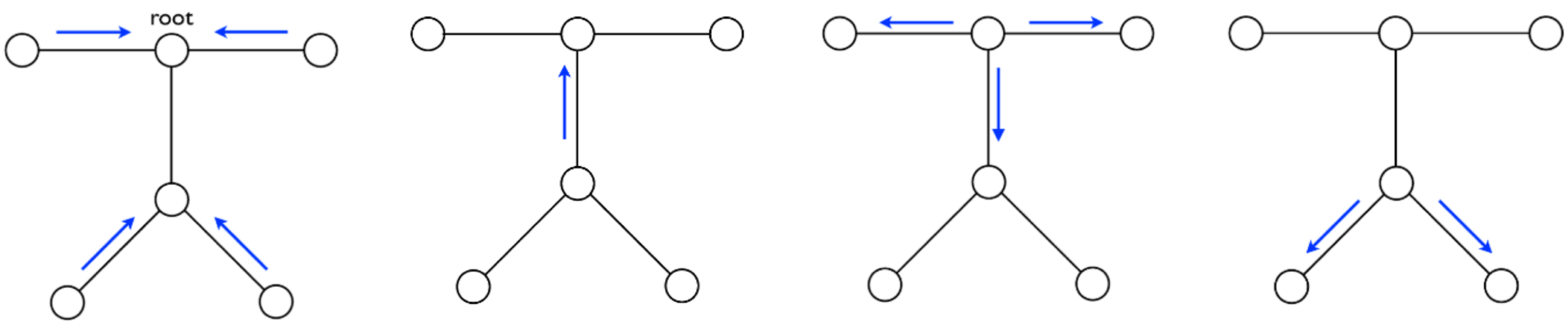
$$p_i(x_i) = \prod_{j \in \eta(i)} m_{ji}(x_i)$$

信念传播算法（BP）

- 消息传递顺序
  - 同时并行传递



  - 深度优先传递

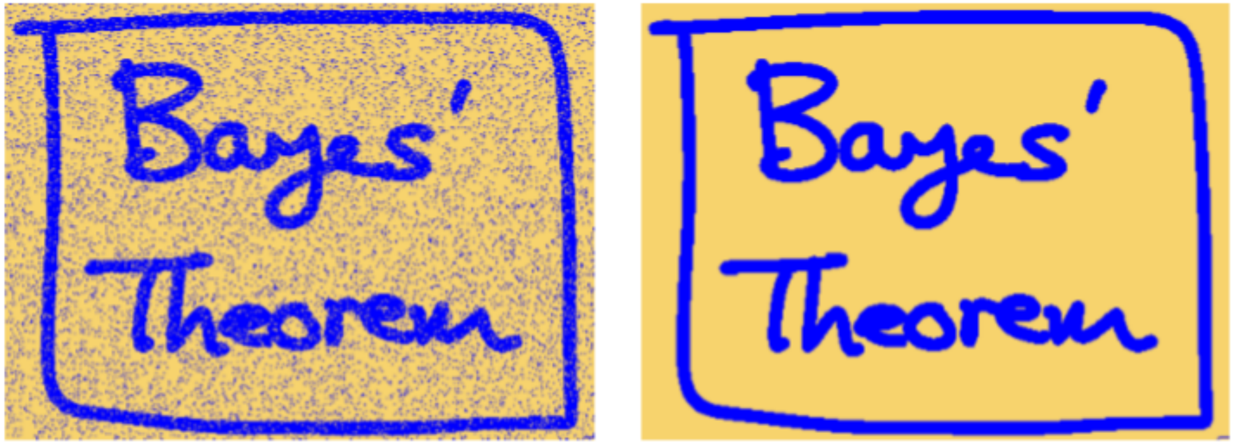


概率图模型（Probabilistic Graphical Models）

- 概率图模型概述
- 概率有向图模型（贝叶斯网络）
  - 因子分解
  - 条件独立性
- 概率无向图模型（马尔可夫随机场）
  - 因子分解
  - 条件独立性
- 信念传播算法
- 应用举例

马尔可夫随机场应用举例

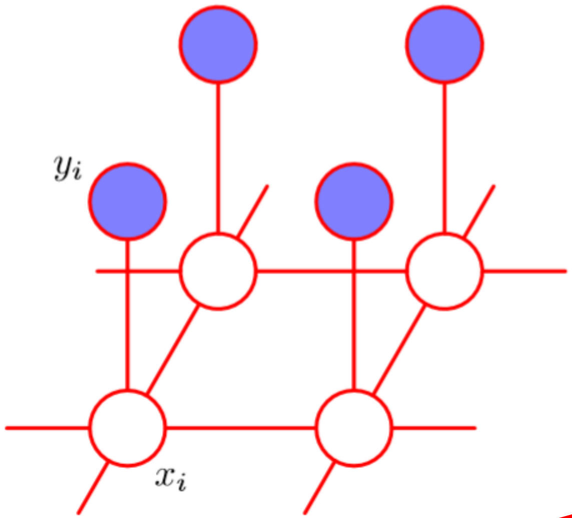
□ 图像去噪



- 带噪图像：$y_i \in \{-1, +1\}$，$i = 1, \ldots, D$覆盖所有像素
  - ✓ 由无噪声图像，以一个较小概率随机翻转像素值符号得到
- 未知的无噪声图像：$x_i \in \{-1, +1\}$
- 目标：给定带有噪声的图像，恢复原始的无噪声图像

马尔可夫随机场

□ 图像去噪



能量函数

两种类型的团

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

- $-\beta x_i x_j$
  - ✓ x_i和x_j符号相同时，具有较低的能量（即较高的概率）
  - ✓ x_i和x_j符号相反时，具有较高的能量（即较低的概率）
- $-\eta x_i y_i$
  - ✓ x_i和y_i符号相同时，具有较低的能量
  - ✓ x_i和y_i符号相反时，具有较高的能量

马尔可夫随机场

图像去噪

联合概率分布

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

固定噪声图像对应的观测值y，图像去噪即求解x，使得条件概率分布$p(x|y)$最大

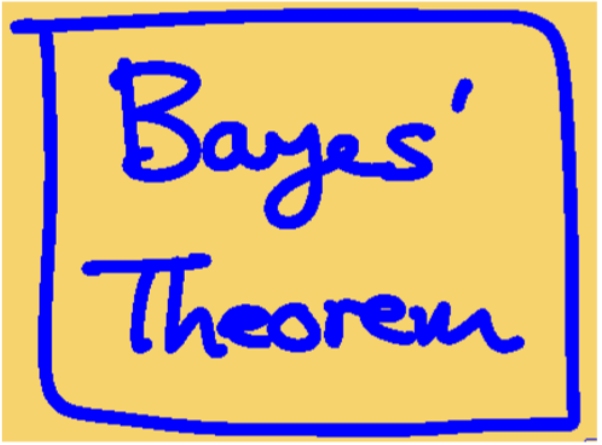
可使用迭代条件模式（ICM）算法，求解获得无噪图像x

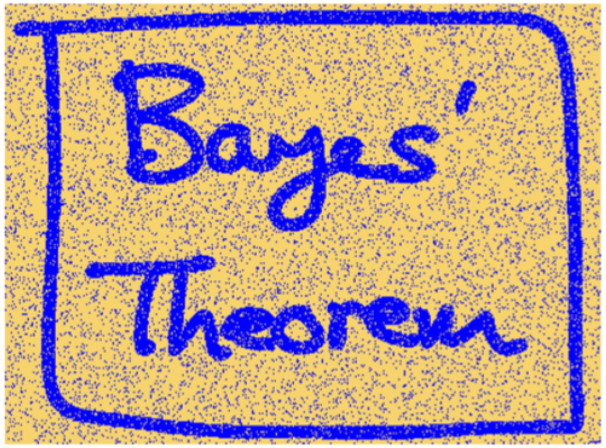
- 对所有像素，初始化变量$x_i=y_i$
- 每次取一个x_j结点，保持其他所有结点变量固定，计算x_j两个可能状态$x_j=+1$和$x_j=-1$的总能量
- 将x_j设置为能量较低的状态
- 对其他结点重复更新过程，直到满足某个合适的停止条件

马尔可夫随机场

- 图像去噪

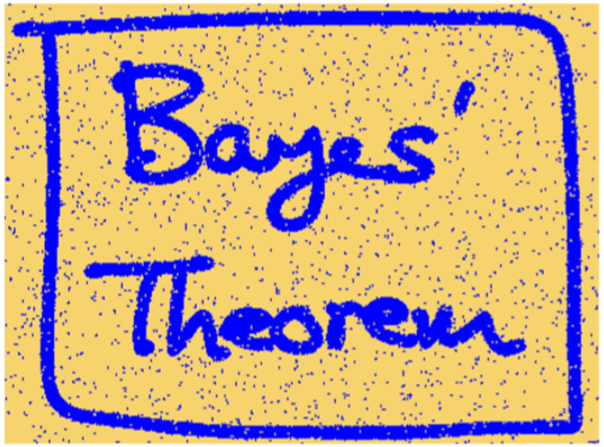
原始二值图像

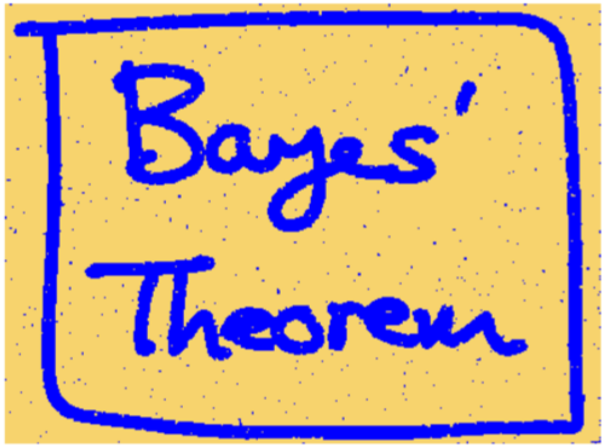




带噪图像

迭代条件模型
恢复的结果





图割算法恢
复的结果