第十一章：基于概率图模型的图像分析

中国科学技术大学
电子工程与信息科学系

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
  - 求和法则：$P(x_1) = \sum_{X_2 \ldots Xn} P(x_1, \cdots, x_N)$
  - 乘积法则：$P(x_1, \cdots, x_N) = P(x_1) \prod_{i=2}^{N} P(x_i | x_1, \cdots, x_{i-1})$
- 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
- 便于分析随机变量之间的各种性质，如条件独立性
- 将复杂计算（如推断、学习问题）表示成图操作

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

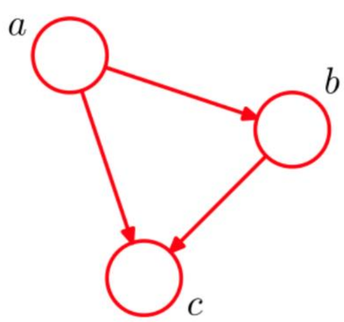
- 概率有向图模型（贝叶斯网络）
  - 因子分解
  - 条件独立性
  - 应用举例：物体识别
- 概率无向图模型（马尔可夫随机场）
  - 条件独立性（马尔可夫性）
  - 因子分解
  - 应用举例：图像去噪
- 条件随机场
  - 定义与形式
  - 概率计算问题
  - 预测算法
  - 应用举例：语义分割

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$p(a,b,c)$

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

- 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

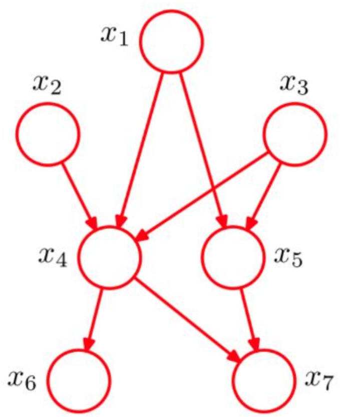
概率有向图模型

- N个随机变量的联合概率分布

$$p(x_1,\cdots,x_N)=p(x_N|x_1,\cdots,x_{N-1})\cdots p(x_2|x_1)p(x_1)$$

  - 任意两个结点间均有边连接

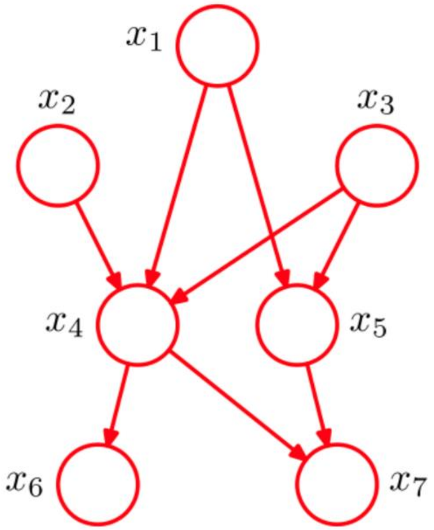
- 缺失的边有意义



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

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



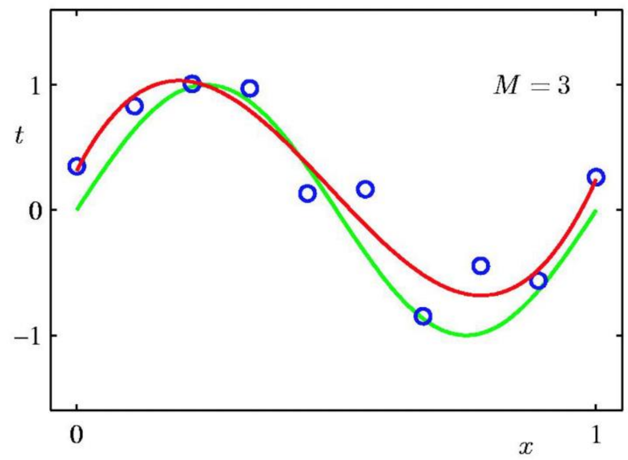
$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i|\mathrm{pa}_i)$$

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

概率有向图模型举例

- 贝叶斯多项式回归
  - 给定输入样本集$\mathbf{X} = \{x_1, x_2, \cdots, x_N\}$ 和对应的目标值$\mathbf{t} = \{t_1, t_2, \cdots, t_N\}$



  - 多项式回归模型

$$y(x, \mathbf{w}) = \sum_{j=0}^{M} w_j x^j$$

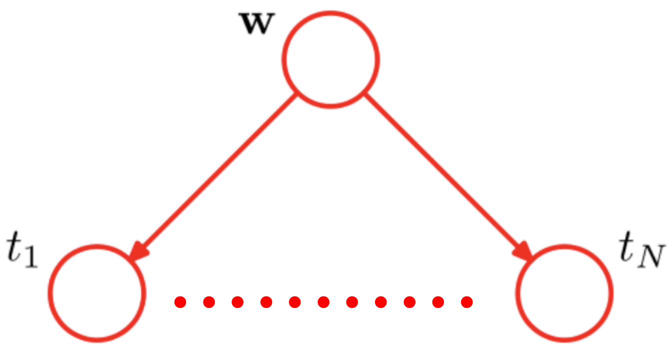
  - 参数**w**和目标**t**的联合概率分布（给定**X** 条件下）

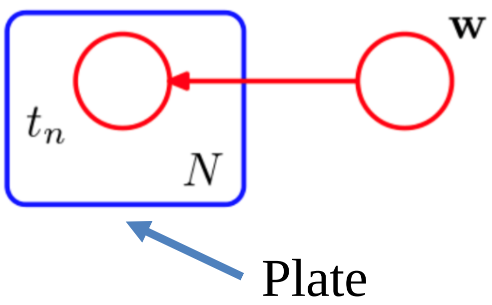
$$p(\mathbf{t}, \mathbf{w}|\mathbf{X}) = \underbrace{p(\mathbf{w})}_{\text{Prior term}} \underbrace{\prod_{\mathrm{n}=1}^{N} p(t_n|y(x_n, \mathbf{w}))}_{\text{Likelihood term}}$$

概率有向图模型举例

$$p(\mathbf{t},\mathbf{w}|\mathbf{X}) = p(\mathbf{w})\prod_{n=1}^{N} p(t_n|y(x_n,\mathbf{w}))$$

□ 概率图模型表示

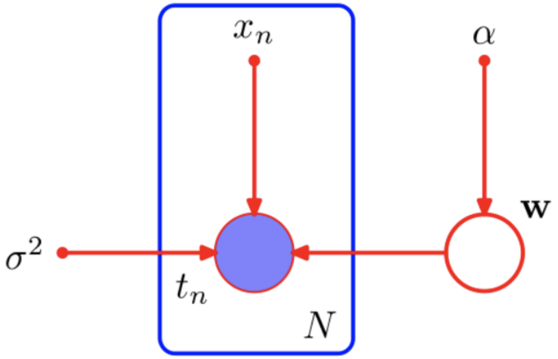




- 紧凑表示（盘式记法）：将由相同机制生成的多个变量放在一个方框（盘）内，并在其中标出变量重复出现的个数 N

概率有向图模型举例

$$p(\mathbf{t},\mathbf{w}|\mathbf{X},\alpha,\sigma^2)=p(\mathbf{w}|\alpha)\prod_{n=1}^{N}p(t_n|x_n,\mathbf{w},\sigma^2)$$



$$p(\mathbf{w}|\alpha)=\mathcal{N}(\mathbf{w}|\mathbf{0},\alpha\boldsymbol{I})$$

$$p(t_n|x_n,\mathbf{w},\sigma^2)=\mathcal{N}(t_n|y(x_n,\mathbf{w}),\sigma^2)$$

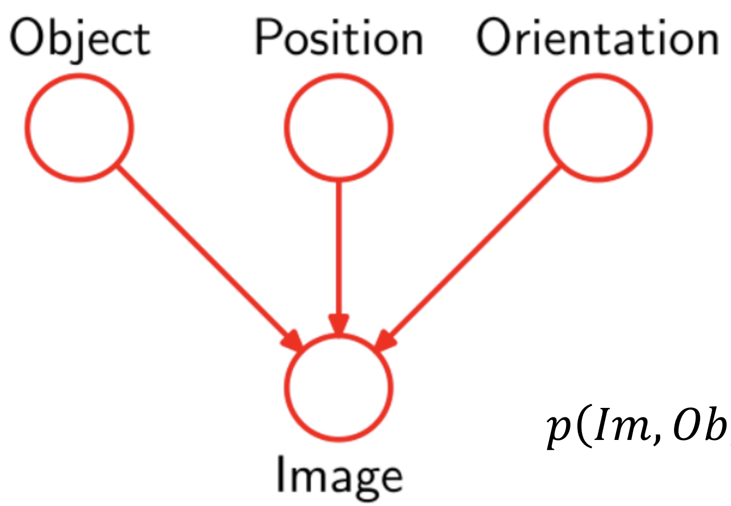
$$y(x,\mathbf{w})=\sum_{j=0}^{M}w_jx^j$$

- 实心小圆表示确定性参数
- 阴影标注已知的、能观察到的变量

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

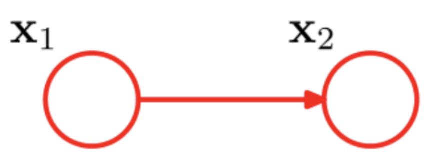


- 物体、位置、方向满足相互独立的先验分布
- 图像的分布以物体、位置、方向为条件

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

离散随机变量

- 考虑离散型随机变量$\mathbf{x}_1, \mathbf{x}_2$，均有$k$种取值，采用1-of-K编码

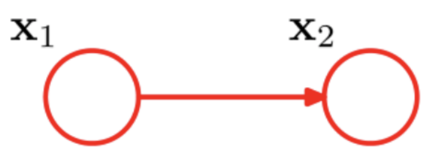


$$p(\mathbf{x}_1, \mathbf{x}_2 | \boldsymbol{\mu}) = \prod_{k=1}^{K} \prod_{l=1}^{K} \mu_{kl}^{x_{1k} x_{2l}}$$

  - x_{1k}表示 $\mathbf{x}_1$的第k个分量，x_{2l}类似，μ_{kl}为$x_{1k} = 1$且$x_{2l} = 1$的概率
  - 该分布包含$k^2 - 1$个参数
- 一般地，M个离散型随机变量的联合分布（对应全连接图）包含 $k^M - 1$个参数，即参数量随变量数M指数增长

离散随机变量

- 一般的两个随机变量的联合分布： k^2-1个参数



$$p(\mathbf{x}_1,\mathbf{x}_2|\boldsymbol{\mu})=\prod_{k=1}^{K}\prod_{l=1}^{K}\mu_{kl}^{x_{1k}x_{2l}}$$

- 两个相互独立的随机变量的联合分布：$2(k-1)$ 个参数

$$p(\mathbf{x}_1,\mathbf{x}_2|\boldsymbol{\mu})=\prod_{k=1}^{K}\mu_{1k}^{x_{1k}}\prod_{l=1}^{K}\mu_{2l}^{x_{2l}}$$

- 通常概率图的连接程度处于无连接和全连接之间，使得对应的分布比完全的独立分布更具一般性，而分布的参数量远小于全连接图的参数量。

链式图

- M个结点的Markov链

$$p(\mathbf{x}_1, \cdots, \mathbf{x}_2) = p(\mathbf{x}_1) \prod_{i=2}^{M} p(\mathbf{x}_i | \mathbf{x}_{i-1})$$

  - $p(\mathbf{x}_1)$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

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

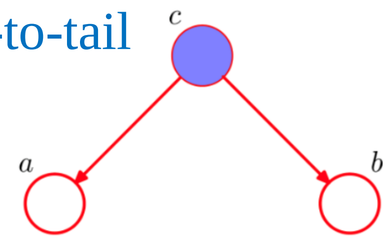
$$\begin{aligned} p(a,b|c) &= p(a|b,c)p(b|c) \\ &= p(a|c)p(b|c) \end{aligned}$$

  - 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

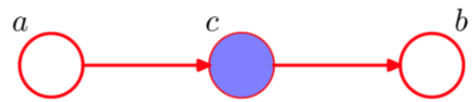
tail-to-tail



$$a \perp b|c$$

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

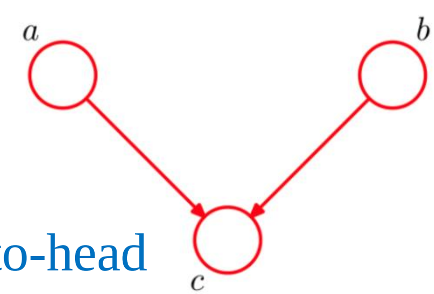
head-to-tail



$$a \perp b|c$$

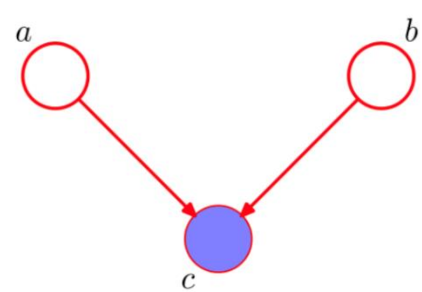
$$p(a,b|c) = \frac{p(a)p(c|a)p(b|c)}{p(c)} = p(a|c)p(b|c)$$

head-to-head



$$a \perp b|\emptyset$$

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



$$a \not\perp b|c$$

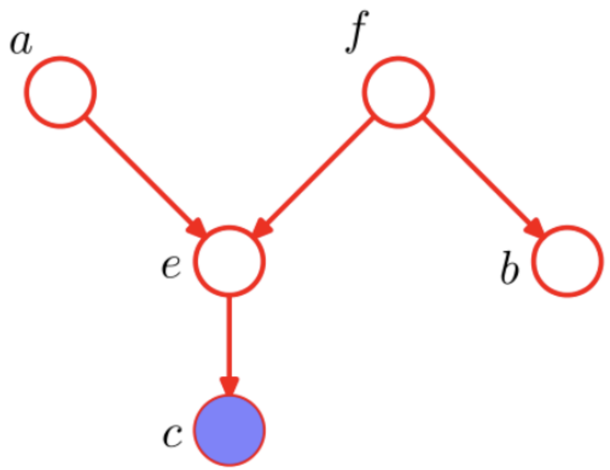
$$p(a,b|c) = \frac{p(a)p(b)p(c|a,b)}{p(c)}$$

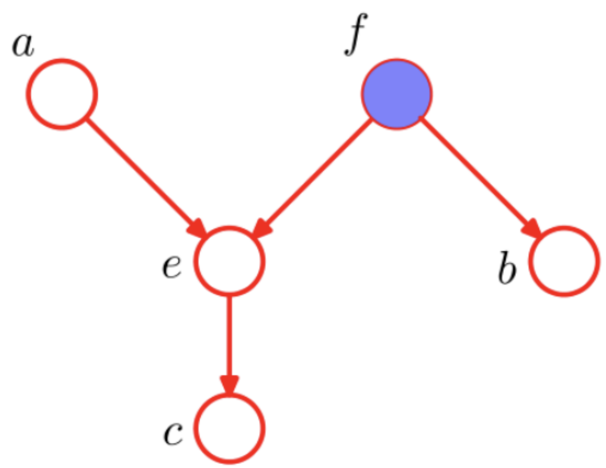
条件独立性

- 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中结点到所有可能路径
  - 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

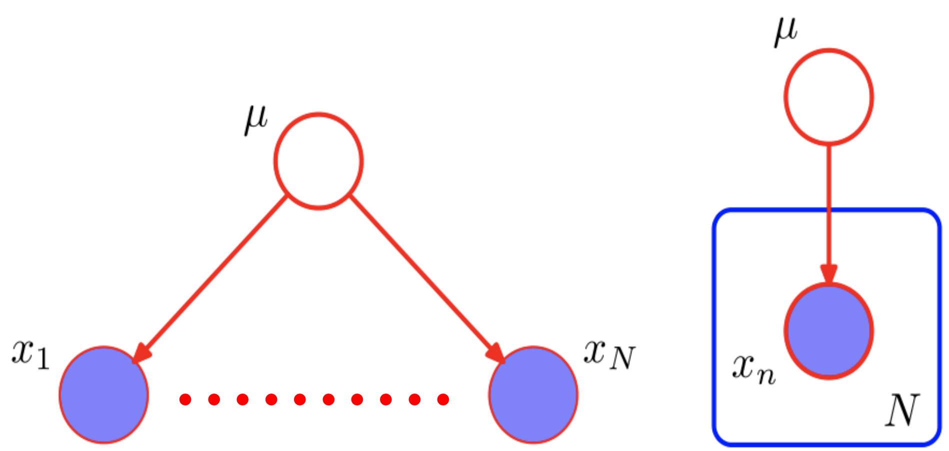
a, b被f阻隔

a, b被e阻隔

$a \perp b | f$成立

条件独立性

- 例子（独立同分布）



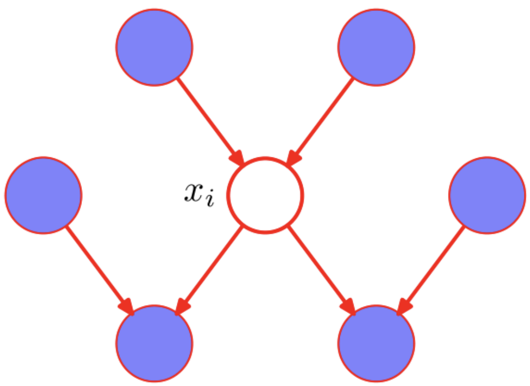
- i.i.d. (independent identically distributed) data

$$p(\mathcal{D}|\mu) = \prod_{n=1}^{N} p(x_n|\mu)$$

$$p(\mathcal{D}) = \int_0^{\infty} p(\mathcal{D}|\mu)p(\mu)\mathrm{d}u \neq \prod_{n=1}^{N} p(x_n)$$

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

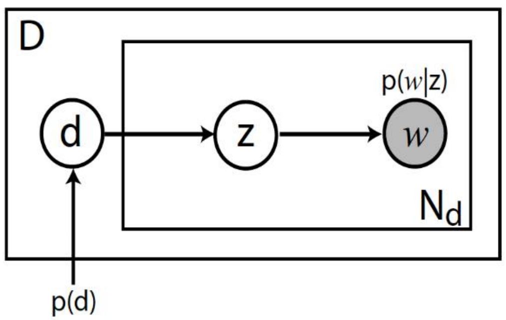


$$p(x_i|x_{\{j\neq i\}}) = \frac{p(x_1,\cdots,x_M)}{\int p(x_1,\cdots,x_M)\mathrm{d}x_i} = \frac{\prod_k p(x_k|\mathrm{pa}_k)}{\int \prod_k p(x_k|\mathrm{pa}_k)\,\mathrm{d}x_i}$$

不含有x_i的因子上下消掉

概率有向图应用举例

- 概率潜在语义分析(probabilistic latent semantic analysis, pLSA)



$$L = \prod_{d=1}^{D} \prod_{w=1}^{W} p(w,d)^{n(w,d)}$$

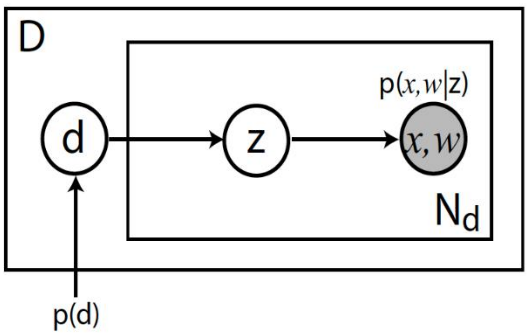
$$\begin{aligned} p(w,d) &= p(d)p(w|d) \\ &= p(d)\textstyle\sum_z p(w,z|d) \\ &= p(d)\textstyle\sum_z p(z|d)p(w|z) \end{aligned}$$

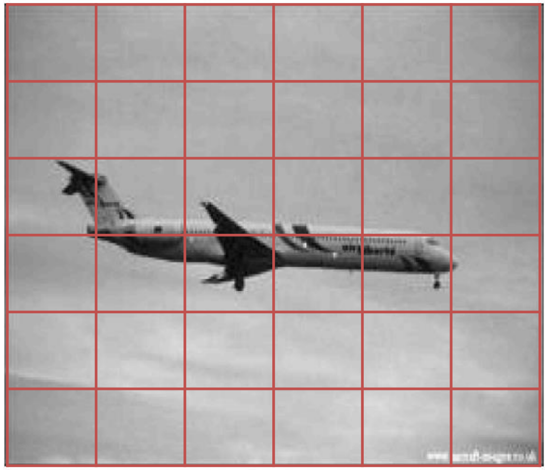
	doc 1	doc 2	doc 3	doc 4
word 1	2	2	4	3
word 2	2	1	5	3
word 3	1	1	2	0
word 4	0	1	2	1

R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning Object Categories from Google's Image Search, in ICCV 2005.

概率有向图应用举例

- ABS-pLSA (Absolute position pLSA)



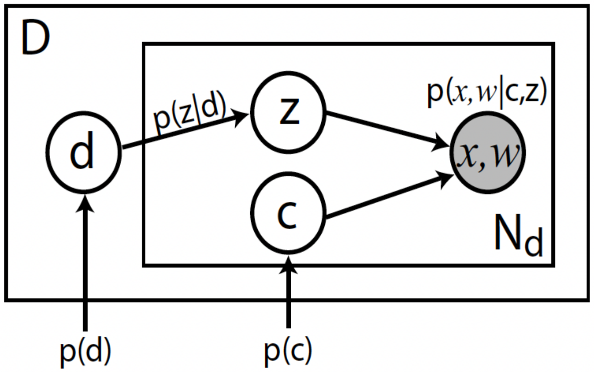


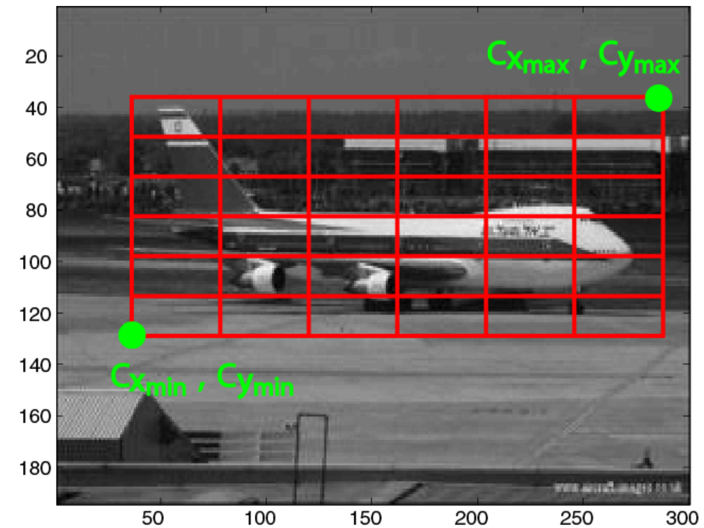
$$p(w,x,d) = p(d)\sum_{z} p(z|d)p(w,x|z)$$

R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning Object Categories from Google's Image Search, in ICCV 2005.

概率有向图应用举例

- TSI-pLSA (Translation and Scale invariant pLSA)

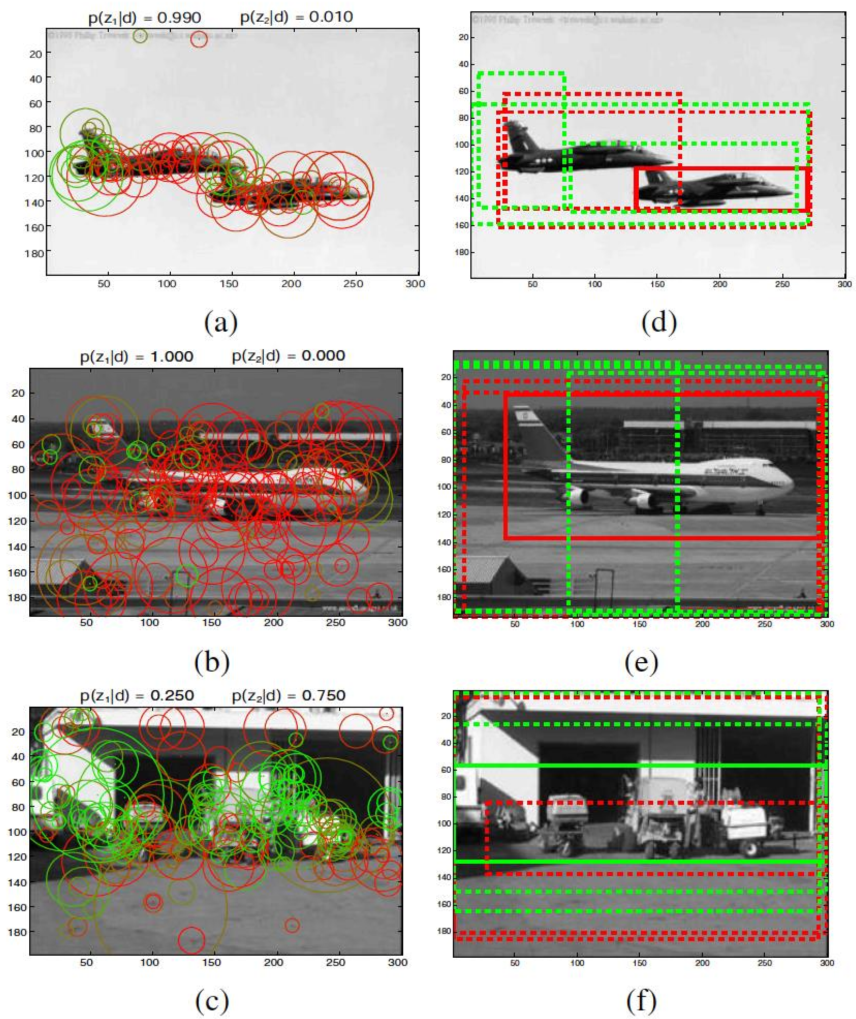




$$p(w,x|z) = \sum_c p(w,x,c|z)$$
$$= \sum_c p(w,x|c,z)p(c)$$

R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning Object Categories from Google's Image Search, in ICCV 2005.

概率有向图应用举例



R. Fergus, L. Fei-Fei, P. Perona, A. Zisserman, Learning Object Categories from Google's Image Search, in ICCV 2005.

概率无向图模型（马尔可夫随机场）

- 模型定义(马尔可夫性)
- 概率无向图模型的因子分解
- 例子：图像去噪

概率无向图模型

- 模型定义
  - 用无向图$G = (V, E)$表示联合概率分布$P(Y)$
  - 无向图表示的随机变量之间存在的性质
    - ✓ 成对马尔可夫性（pairwise Markov property）
    - ✓ 局部马尔可夫性（local Markov property）
    - ✓ 全局马尔可夫性（global Markov property）

模型定义

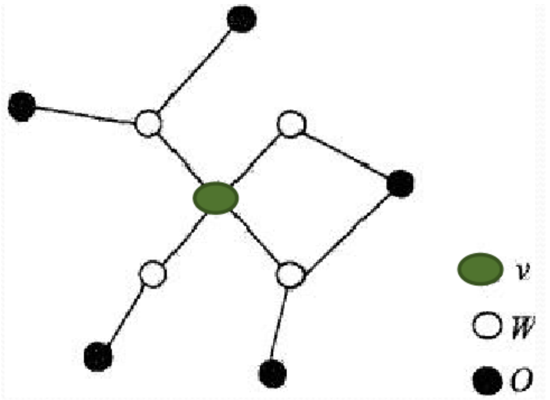
- 成对马尔可夫性（Pairwise Markov property）
  - 设u和v是无向图G中任意两个没有边连接的结点，结点u和v分别对应随机变量Y_u和Y_v
  - 其他所有结点为O，对应的随机变量组是Y_O
  - 定义：给定随机变量组Y_O的条件下，随机变量Y_u和Y_v是条件独立的

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

模型定义

- 局部马尔可夫性（Local Markov property）
  - $v \in V$是任意结点
  - 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Y
    - ✓ v：Y_v，W：Y_W，O：Y_O
  - 定义：给定随机变量组Y_W的条件下随机变量Y_v与随机变量Y_O是独立的

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



  - 在$P(Y_O|Y_W) > 0$时，等价于

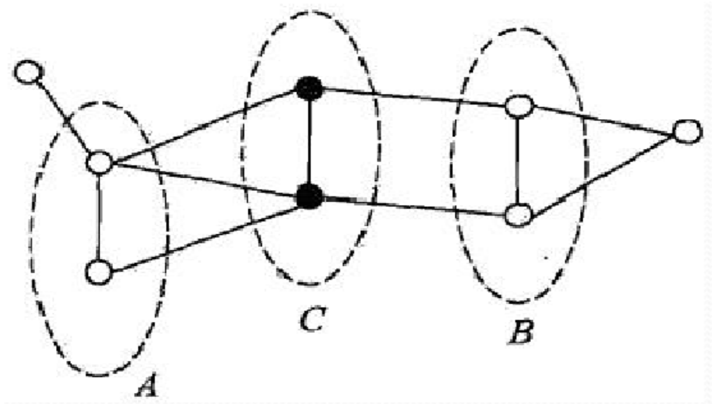
$$P(Y_v|Y_W, Y_O) = P(Y_v|Y_W)$$

  - 马尔可夫毯：与一个结点相邻的结点构成的集合

模型定义

- 全局马尔可夫性（Global Markov property）
  - 结点集合A，B是在无向图G中被结点集合C分开的任意结点集合
  - A，B，C对应的随机变量组分别为Y_A，Y_B，Y_C
  - 定义：给定随机变量组Y_C的条件下随机变量组Y_A与随机变量组Y_B是条件独立的

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

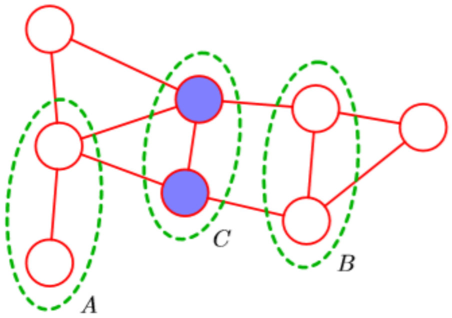


模型定义

- 全局马尔可夫性

在无向图中，有三个结点集合A、B、C

- 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 → 条件独立
- 把集合C中的结点以及与这些结点相连的链接全部删除，考察是否存在一条从A中任意结点到B中任意结点的路径，如果没有，那么条件独立性质成立



C为条件，A与B条件独立

模型定义

- 概率无向图模型
  - 设有联合概率分布 $P(Y)$，由无向图 $G=(V,E)$ 表示，在图 G 中，结点表示随机变量，边表示随机变量之间的依赖关系
  - 如果联合概率分布$P(Y)$满足成对、局部或全局马尔可夫性，就称此联合概率分布为概率无向图模型（probabilistic undirected graphical model），或马尔可夫随机场（Markov random field）
  - 联合概率分布的表示：因子分解

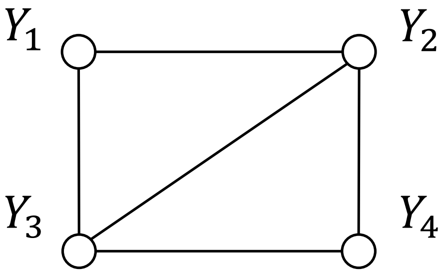
概率无向图模型（马尔可夫随机场）

- 模型定义
- 概率无向图模型因子分解
- 例子：图像去噪

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）

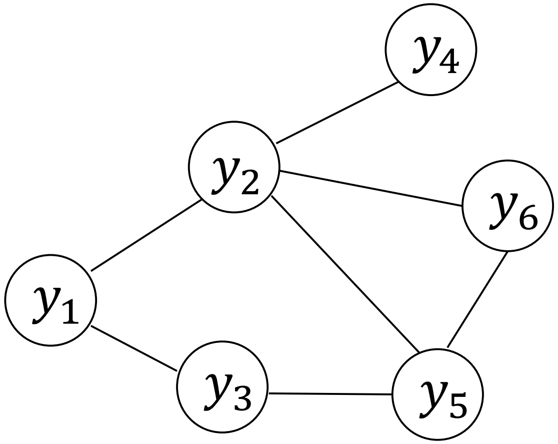
- ✓ 划分函数的计算复杂度是指数级的
- ✓ 对于参数学习来说，划分函数是必要的；对于局部条件概率分布的计算，划分函数不是必要的

概率无向图模型的因子分解

- 联合概率分布定义

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

- 举例
  - 图模型

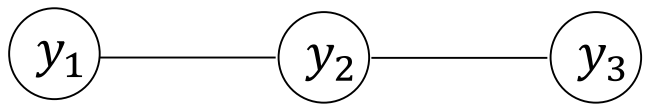


  - 联合概率分布

$$P(Y) = \frac{1}{Z}\psi_{12}(y_1,y_2)\psi_{13}(y_1,y_3)\ \psi_{24}(y_2,y_4)\ \psi_{35}(y_3,y_5)\ \psi_{256}(y_2,y_5,y_6)$$

概率无向图模型的因子分解

- 势函数$\psi_C(Y_C)$的作用是定量刻画变量集Y_C中变量的相关关系，应为非负函数，且在所偏好的变量取值上有较大的函数值



- 上图中，假定变量均为二值变量，若定义势函数

$$\psi_{12}(y_1, y_2) = \begin{cases} 1.5, & if\ y_1 = y_2 \\ 0.1, & otherwise \end{cases}$$

$$\psi_{23}(y_2, y_3) = \begin{cases} 0.2, & if\ y_2 = y_3 \\ 1.3, & otherwise \end{cases}$$

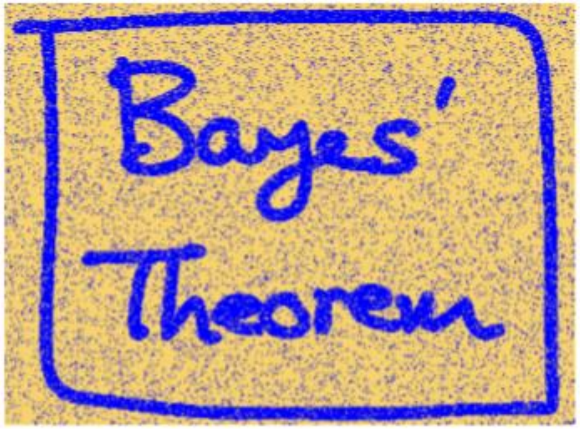
- 说明模型偏好y_1与y_2有相同的取值，y_2与y_3有不同的取值；换言之，y_1与y_2正相关，y_2与y_3负相关。所以令y_1与y_2相同且y_2与y_3不同的变量值指派将有较高的联合概率

概率无向图模型（马尔可夫随机场）

- 模型定义
- 概率无向图模型因子分解
- 例子：图像去噪

马尔可夫随机场

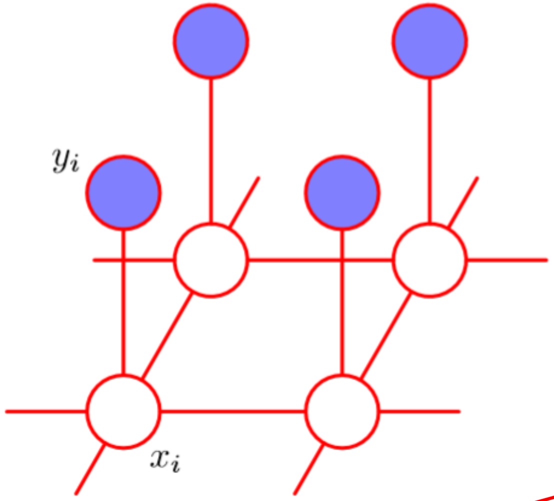
图像去噪



- 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

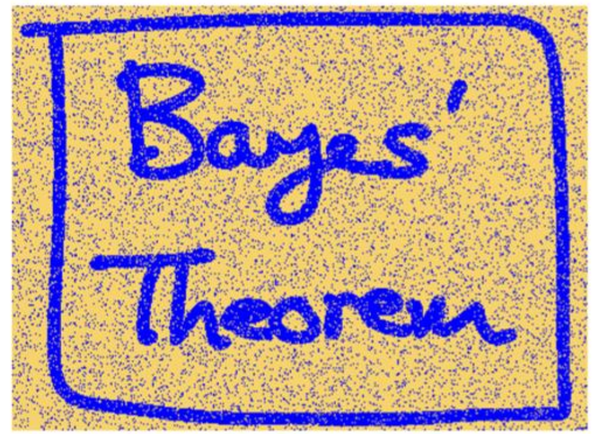
可使用迭代条件模式（ICM）算法，求解获得无噪图像x

- 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

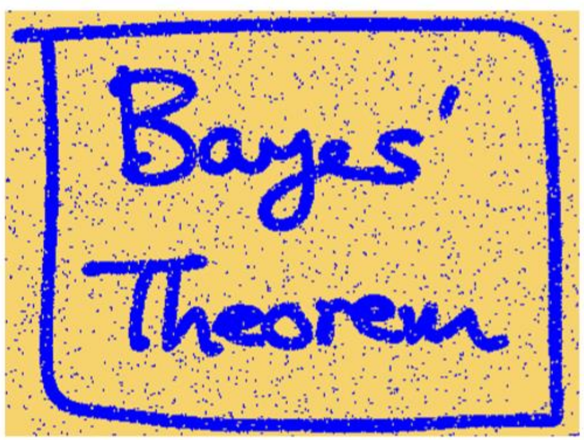
□ 图像去噪

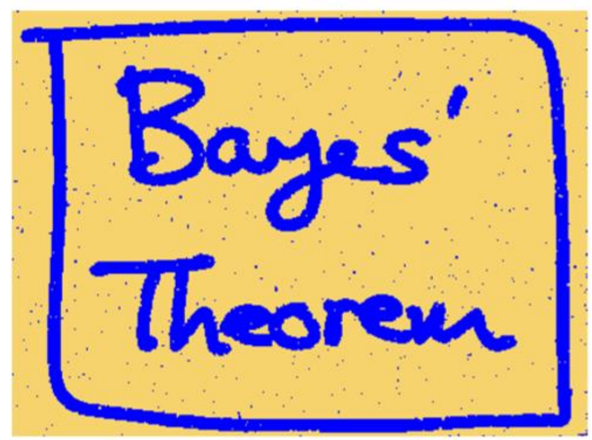
原始二值图像



带噪图像

迭代条件模型
恢复的结果

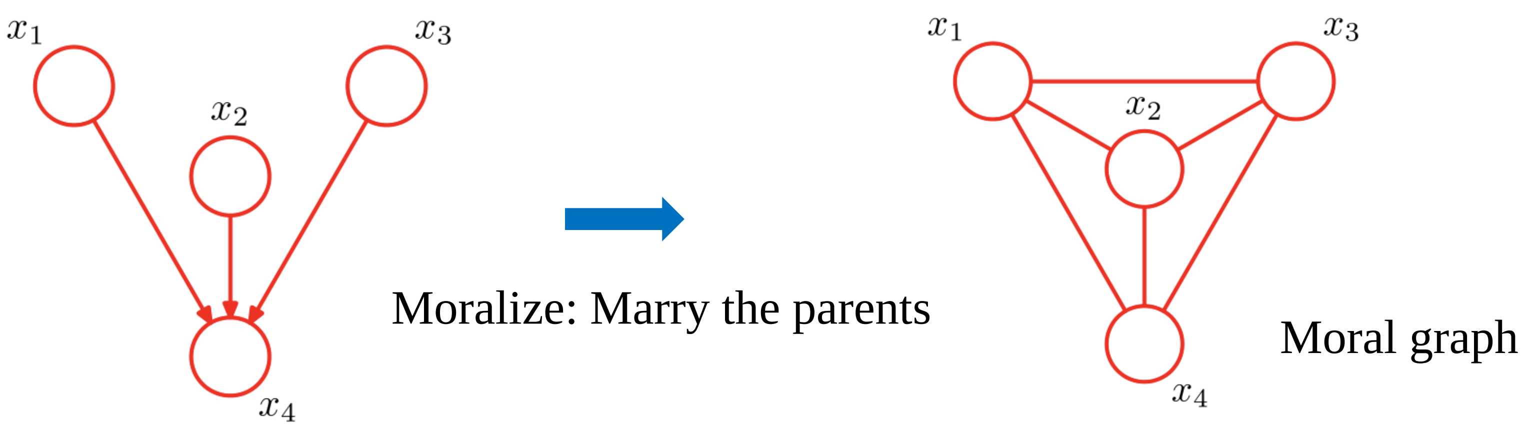




图割算法恢
复的结果

有向图和无向图模型的关系

□ 有向图模型与无向图模型之间的转换



$$p(\mathbf{x}) = p(x_1)p(x_2)p(x_3)p(x_4|x_1,x_2,x_3) \qquad p(\mathbf{x}) = \frac{1}{Z}\psi(x_1,x_2,x_3,x_4)$$

- 为了表示有向图中的高阶关系，无向图中需要包含相应结点的团
- 有向图中的条件独立关系可能丢失

概率图模型

- 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）
  - 条件随机场

条件随机场

- □ 条件随机场定义与形式
- □ 条件随机场概率计算问题
- □ 条件随机场预测算法
- □ 例子：图像语义分割

条件随机场定义

- 条件随机场
  - 给定随机变量X条件下，随机变量Y的马尔可夫随机场
- 条件随机场的定义
  - 设X与Y是随机变量，$P(Y|X)$是在给定X的条件下Y的条件概率分布，若随机变量Y构成一个由无向图$G=(V,E)$表示的马尔可夫随机场，即满足马尔可夫性

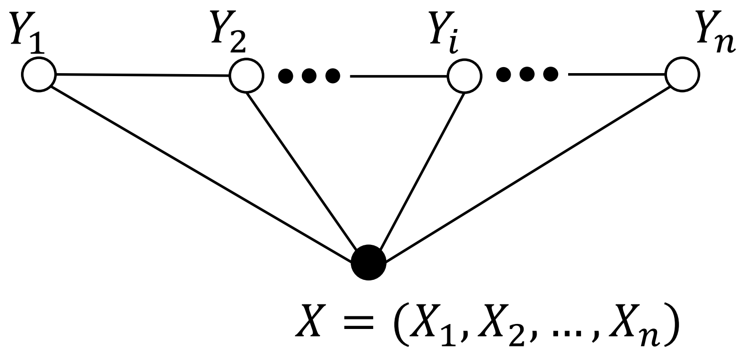
$$P(Y_v|X,Y_w,w\neq v)=P(Y_v|X,Y_w,w\sim v)$$

对任意结点v成立，则称概率分布$P(Y|X)$为条件随机场，式中$w\neq v$表示结点v以外的所有结点，$w\sim v$表示在图$G=(V,E)$中与结点v有边连接的所有结点w。

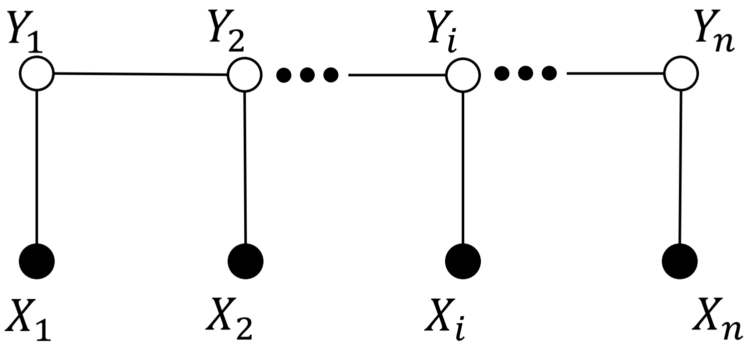
条件随机场定义

- 线性链条件随机场（linear chain conditional random field）
  - 定义在线性链上的特殊的条件随机场

$$G = (V = \{1,2,\ldots,n\}, E = \{(i,i+1)\}), i = 1,2,\ldots,n-1$$



线性链条件随机场



X和Y有相同的图结构的线性链条件随机场

  - 最大团是相邻两个结点的集合
  - 可以用于标注等问题

条件随机场定义

- 线性链条件随机场
  - 设$X=(X_1,X_2,\dots,X_n)$，$Y=(Y_1,Y_2,\dots,Y_n)$均为线性链表示的随机变量序列，若在给定随机变量序列X的条件下，随机变量序列Y的条件概率分布$P(Y|X)$构成条件随机场，即满足马尔可夫性

$$P(Y_i|X,Y_1,\dots,Y_{i-1},Y_{i+1},\dots,Y_n)=P(Y_i|X,Y_{i-1},Y_{i+1})$$

$$i=1,2,\dots,n \text{（在}i=1\text{和}n\text{时只考虑单边）}$$

  称$P(Y|X)$为线性链条件随机场

  - 在标注问题中，X表示输入观测序列，Y表示对应的输出标记序列或状态序列

条件随机场参数化形式

□ 线性链条件随机场的参数化形式

- 设$P(Y|X)$为线性链条件随机场，则在随机变量X取值为x的条件下，随机变量Y取值为y的条件概率具有如下形式：

$$P(y|x) = \frac{1}{Z(x)} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right)$$

$$Z(x) = \sum_{y} \exp\left(\sum_{i,k} \lambda_k t_k(y_{i-1}, y_i, x, i) + \sum_{i,l} \mu_l s_l(y_i, x, i) \right)$$

- ✓ $Z(x)$：规范化因子
- ✓ 转移特征：t_k是定义在边上的特征函数，依赖于前一个和当前位置
- ✓ 状态特征：s_l是定义在结点上的特征函数，依赖于当前位置
- ✓ λ_k和μ_l：t_k和s_l对应的权值

条件随机场的简化形式

- □ 条件随机场的简化形式

  条件随机场中同一特征在各个位置都有定义，可以对同一个特征在各个位置求和，将局部特征函数转化为一个全局特征函数，这样就可以将条件随机场写成权值向量和特征向量的内积形式，即条件随机场的简化形式

简化步骤：

- □ 首先将转移特征和状态特征及其权值用统一的符号表示。设有K_1个转移特征，K_2个状态特征，$K = K_1 + K_2$，记

$$f_k(y_{i-1}, y_i, x, i) = \begin{cases} t_k(y_{i-1}, y_i, x, i), & k = 1,2, \dots, K_1 \\ s_l(y_i, x, i), & k = K_1 + l; l = 1,2, \dots, K_2 \end{cases}$$

条件随机场的简化形式

- 对转移特征与状态特征在各个位置i求和

$$f_k(y,x)=\sum_{i=1}^{n} f_k(y_{i-1},y_i,x,i)\,,\qquad k=1,2,\dots,K$$

- 权值：

$$w_k=\begin{cases}\lambda_k, & k=1,2,\dots,K_1\\ \mu_l, & k=K_1+l;\, l=1,2,\dots,K_2\end{cases}$$

- 条件随机场可表示为

$$P(y|x)=\frac{1}{Z(x)}\exp\sum_{k=1}^{K} w_k f_k(y,x)\qquad Z(x)=\sum_{y}\exp\sum_{k=1}^{K} w_k f_k(y,x)$$

条件随机场的简化形式

- w表示权值向量

$$w = (w_1, w_2, \dots, w_K)^T$$

- 以$F(y,x)$表示全局特征向量，即

$$F(y,x) = \big(f_1(y,x), f_2(y,x), \dots, f_K(y,x)\big)^T$$

- 条件随机场内积形式

$$P_w(y|x) = \frac{\exp(w \cdot F(y,x))}{Z_w(x)}$$

$$Z_w(x) = \sum_y \exp(w \cdot F(y,x))$$

条件随机场的矩阵形式

- 对线性链条件随机场

$$P_w(y|x)=\frac{1}{Z(x)}\exp\left(\sum_i\sum_k w_k f_k(y_{i-1},y_i,x,i)\right)$$

引进特殊的起点和终点状态标记$Y_0=start$，$Y_{n+1}=stop$，这时$P_w(y|x)$可以通过矩阵形式表示

- 对观测序列x的每一个位置$i=1,2,\dots,n+1$，定义一个m阶矩阵$M_i(x)\in\mathcal{R}^{m\times m}$（$m$是标记$Y_i$取值的个数）

$$M_i(x)=[M_i(y_{i-1},y_i|x)]$$

$$M_i(y_{i-1},y_i|x)=\exp(W_i(y_{i-1},y_i|x))$$

$$W_i(y_{i-1},y_i|x)=\sum_{k=1}^{K}w_k f_k(y_{i-1},y_i,x,i)$$

条件随机场的矩阵形式

- 线性链条件随机场

$$P_w(y|x) = \frac{1}{Z(x)} \exp\left(\sum_i \sum_k w_k f_k(y_{i-1}, y_i, x, i)\right)$$

$M_1(y_0, y_1|x)$

$M_2(y_1, y_2|x)$

$$= \frac{1}{Z(x)} \exp\left(\sum_k w_k f_k(y_0, y_1, x, 1)\right) \exp\left(\sum_k w_k f_k(y_1, y_2, x, 2)\right) \cdots$$

$$\exp\left(\sum_k w_k f_k(y_n, y_{n+1}, x, n+1)\right)$$

$M_{n+1}(y_n, y_{n+1}|x)$

条件随机场的矩阵形式

- 给定观测序列x，标记序列y的非规范化概率可以通过$n+1$个适当的矩阵元素的乘积表示：

$$\prod_{i=1}^{n+1} M_i(y_{i-1}, y_i|x)$$

- 条件概率$P_w(y|x)$：

$$P_w(y|x) = \frac{1}{Z_w(x)} \prod_{i=1}^{n+1} M_i(y_{i-1}, y_i|x)$$

- $Z_w(x)$为规范化因子，可由$n+1$个矩阵相乘得到

$$Z_w(x) = [M_1(x) M_2(x) \cdots M_{n+1}(x)]_{\text{start,stop}}$$

条件随机场的矩阵形式

- 例 1： 给定一个线性链条件随机场，观测序列 x ， 状态序列y，$n=3$，标记 $y_i \in \{1,2\}, i=1,2,3$，假设 $y_0=\text{start}=1, y_4=\text{stop}=1$，各个位置的随机矩阵$M_1(x), M_2(x), M_3(x), M_4(x)$分别是

$$M_1(x)=\begin{bmatrix} a_{01} & a_{02} \\ 0 & 0 \end{bmatrix},\ M_2(x)=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix},$$

$$M_3(x)=\begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix},\ M_4(x)=\begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix},$$

试求状态序列 y 以 start 为起点 stop 为终点所有路径的非规范化概率及规范化因子。

条件随机场

- 条件随机场定义与形式
- 条件随机场概率计算问题
- 条件随机场预测算法
- 例子：图像语义分割

条件随机场的概率计算问题

- 概率计算问题
  - 给定条件随机场$P(Y|X)$，输入序列x和输出序列y
  - 计算条件概率：$P(Y_i = y_i|x)$，$P(Y_{i-1} = y_{i-1}, Y_i = y_i|x)$
  - 以及相应的数学期望

$P(Y_i = y_i|x)$

$$= \sum_{y_1} \cdots \sum_{y_{i-1}} \sum_{y_{i+1}} \cdots \sum_{y_n} \frac{1}{Z_w(x)} \prod_{j=1}^{n+1} M_j(y_{j-1}, y_j|x)$$

$$= \frac{1}{Z_w(x)} \left(\sum_{y_1} \cdots \sum_{y_{i-1}} \prod_{j=1}^{i} M_j(y_{j-1}, y_j|x) \right) \times \left(\sum_{y_{i+1}} \cdots \sum_{y_n} \prod_{j=i+1}^{n+1} M_j(y_{j-1}, y_j|x) \right)$$

$$= \frac{1}{Z_w(x)} \left(\sum_{y_{i-1}} M_i(y_{i-1}, y_i|x) \cdots \sum_{y_2} M_3(y_2, y_3|x) \sum_{y_1} M_1(y_0, y_1|x) M_2(y_1, y_2|x) \right)$$
$$\times \left(\sum_{y_{i+1}} M_{i+1}(y_i, y_{i+1}|x) \cdots \sum_{y_{n-1}} M_{n-1}(y_{n-2}, y_{n-1}|x) \sum_{y_n} M_n(y_{n-1}, y_n|x) M_{n+1}(y_n, y_{n+1}|x) \right)$$

条件随机场的概率计算问题

- 前向-后向算法：前向
  - 对每个指标$i = 0,1,\dots,n+1$，定义前向向量$\alpha_i(x)$
  - 递推公式

$$\alpha_0(y_0|x) = \begin{cases} 1, & y_0 = start \\ 0, & \text{否则} \end{cases}$$

  - 对 $i = 1,2,\dots,n+1$

$$\alpha_i^T(x) = \alpha_{i-1}^T(x) M_i(x)$$

  - $\alpha_i(y_i|x)$表示在位置i的标记是y_i，且到位置i的前部分标记序列的非规范化概率，y_i可取的值m个，所以$\alpha_i(x)$是m维列向量

条件随机场的概率计算问题

- 前向-后向算法：后向
  - 对每个指标$i = 0,1,\dots,n+1$，定义后向向量$\beta_i(x)$
  - 递推公式

$$\beta_{n+1}(y_{n+1}|x) = \begin{cases} 1, & y_{n+1} = stop \\ 0, & \text{否则} \end{cases}$$

  - 对 $i = n, n-1, \dots, 0$

$$\beta_i(x) = M_{i+1}(x)\beta_{i+1}(x)$$

  - $\beta_i(y_i|x)$表示在位置i的标记是y_i，且从位置$i+1$到n的后部分标记序列的非规范化概率
- 由前向-后向向量得

$$Z(x) = \alpha_n^T(x)\mathbf{1} = \mathbf{1}^T\beta_0(x)$$

条件随机场的概率计算问题

- □ 概率计算
  - ■ 按照前向-后向向量的定义，计算标记序列在位置i是标记y_i的条件概率和在位置$i-1$与i是标记y_{i-1}和y_i的条件概率

$$P(Y_i = y_i|x) = \frac{\alpha_i^T(y_i|x)\beta_i(y_i|x)}{Z(x)}$$

$$P(Y_{i-1} = y_{i-1}, Y_i = y_i|x) = \frac{\alpha_{i-1}^T(y_{i-1}|x)M_i(y_{i-1}, y_i)\beta_i(y_i|x)}{Z(x)}$$

$$\boxed{Z(x) = \alpha_n^T(x)\mathbf{1}}$$

条件随机场的概率计算问题

□ 期望值的计算

- 利用前向-后向向量，可以计算特征函数关于联合分布$P(X,Y)$和条件分布$P(Y|X)$的数学期望
- 特征函数f_k关于条件分布$P(Y|X)$的数学期望是：

$$E_{P(Y|X)}[f_k] = \sum_{y} P(y|x) f_k(y,x)$$

$$= \sum_{i=1}^{n+1} \sum_{y_{i-1} y_i} f_k(y_{i-1}, y_i, x, i) \frac{\alpha_{i-1}^T(y_{i-1}|x) M_i(y_{i-1}, y_i) \beta_i(y_i|x)}{Z(x)}$$

$$k = 1,2,\dots,K$$

$$\boxed{Z(x) = \alpha_n^T(x)\mathbf{1}}$$

条件随机场的概率计算问题

- 假设经验分布为$\tilde{P}(X)$特征函数f_k关于联合分布$P(X,Y)$的数学期望是：

$$E_{P(X,Y)}[f_k]=\sum_{x,y}P(x,y)\sum_{i=1}^{n+1}f_k(y_{i-1},y_i,x,i)$$

$$=\sum_{x}\tilde{P}(x)\sum_{y}P(y|x)\sum_{i=1}^{n+1}f_k(y_{i-1},y_i,x,i)$$

$$=\sum_{x}\tilde{P}(x)\sum_{i=1}^{n+1}\sum_{y_{i-1}y_i}f_k(y_{i-1},y_i,x,i)\frac{\alpha_{i-1}^T(y_{i-1}|x)M_i(y_{i-1},y_i)\beta_i(y_i|x)}{Z(x)}$$

$$k=1,2,\dots,K$$

$$\boxed{Z(x)=\alpha_n^T(x)\mathbf{1}}$$

条件随机场

- 条件随机场定义与形式
- 条件随机场概率计算问题
- 条件随机场预测算法
- 例子：图像语义分割

条件随机场的预测算法

- 预测算法：
  - 给定条件随机场$P(Y|X)$和输入序列（观测序列）x
  - 求：条件概率最大的输出序列（标记序列）y^*

- 维特比算法

$$P_w(y|x) = \frac{\exp(w \cdot F(y,x))}{Z_w(x)}$$

$$\begin{aligned} y^* &= \arg\max_y P_w(y|x) \\ &= \arg\max_y \frac{\exp(w \cdot F(y,x))}{Z_w(x)} \\ &= \arg\max_y \exp(w \cdot F(y,x)) \\ &= \arg\max_y w \cdot F(y,x) \end{aligned}$$

非规范化概率最大的最优路径

$$\max_y w \cdot F(y,x)$$

条件随机场的预测算法

- 求非规范化概率最大的最优路径

$$\arg\max_{y} w \cdot F(y, x)$$

$$= \arg\max_{y} \sum_{i=1}^{n} w \cdot F_i(y_{i-1}, y_i, x)$$

其中

$$w = (w_1, w_2, \dots, w_K)^T$$

$$F_i(y_{i-1}, y_i, x) = \left(f_1(y_{i-1}, y_i, x, i), f_2(y_{i-1}, y_i, x, i), \dots, f_K(y_{i-1}, y_i, x, i)\right)^T$$

条件随机场的预测算法

- 维特比算法
  - 首先求出位置1的各个标记$l=1,2,\ldots,m$的非规范化概率

$$\delta_1(l)=w\cdot F_1\,(y_0=start,y_1=l,x),l=1,2,\ldots,m$$

  - 由递推公式，求出到位置i的各个标记$l=1,2,\ldots,m$的非规范化概率的最大值，同时记录最大值路径

$$\delta_i(l)=\max_{1\le j\le m}\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\,,l=1,2,\ldots,m$$

$$\psi_i(l)=\arg\max_{1\le j\le m}\delta_{i-1}(j)+w\cdot F_i(y_{i-1}=j,y_i=l,x)\,,l=1,2,\ldots,m$$

条件随机场的预测算法

- 维特比算法
  - 直到$i=n$时终止，这时求得非规范概率的最大值为

$$\max_{y}\left(w\cdot F(y,x)\right)=\max_{1\le j\le m}\delta_n(j)$$

  - 最优路径的终点

$$y_n^*=\arg\max_{1\le j\le m}\delta_n(j)$$

  - 由此最优路径终点返回

$$y_i^*=\psi_{i+1}(y_{i+1}^*),i=n-1,n-2,\dots,1$$

  - 得最优路径

$$y^*=(y_1^*,y_2^*,..,y_n^*)^T$$

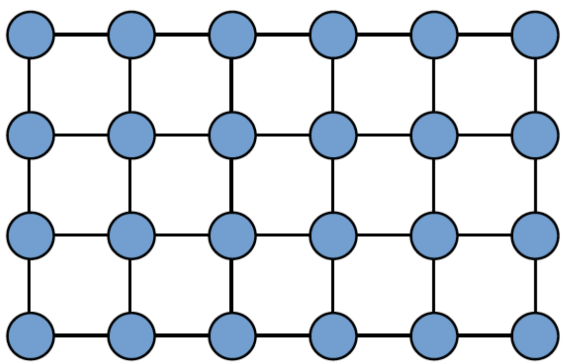
条件随机场

- 条件随机场定义与形式
- 条件随机场概率计算问题
- 条件随机场预测算法
- 例子：图像语义分割

基于条件随机场的图像分割

- 能量函数

$$E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j \in \mathcal{N}_i} \psi_p(x_i, x_j)$$



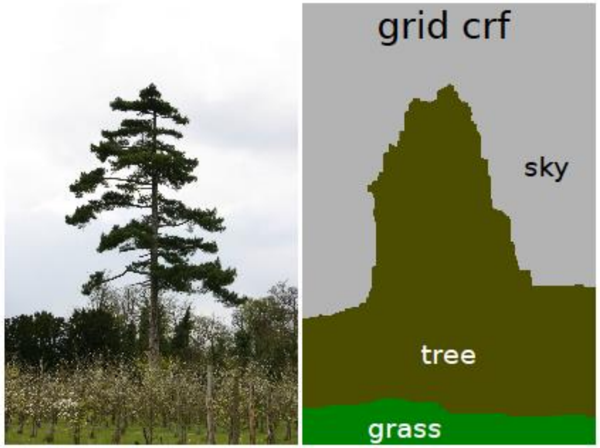
- $\psi_u(x_i)$
  - 来自分类器（像素）
  - 利用颜色、纹理、位置等特征
- $\psi_p(x_i, x_j)$
  - 考虑相邻像素

$$\psi_p(x_i, x_j) = 1_{[x_i \neq x_j]} \left(w^{(1)} \exp\left(-\frac{|\mathrm{I}_i - \mathrm{I}_j|^2}{2\theta_\beta^2} \right) + w^{(2)} \right)$$

基于条件随机场的图像分割

- 能量函数

$$E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j \in \mathcal{N}_i} \psi_p(x_i, x_j)$$

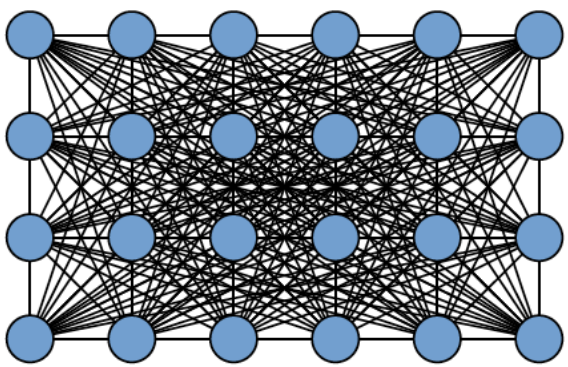


- 推断高效
- 仅考虑了局部关系
- 物体边界过度平滑

全连接条件随机场

□ 能量函数

$$E(x)=\sum_i \psi_u(x_i)+\sum_i\sum_{j>i}\psi_p(x_i,x_j)$$



□ 每个结点和所有其他结点相连

□ 不同的连接具有不同的权重

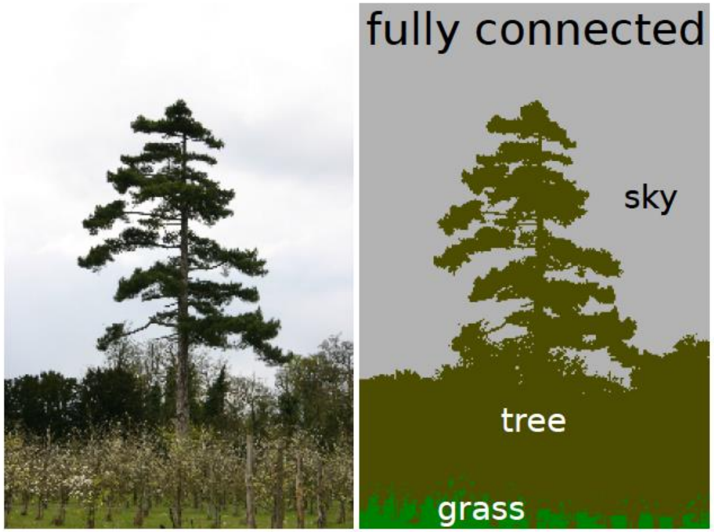
□ 体现远距离相关性

P. Krähenbühl and V. Koltun, Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials, in NIPS 2011.

全连接条件随机场

- 能量函数

$$E(x) = \sum_i \psi_u(x_i) + \sum_i \sum_{j>i} \psi_p(x_i, x_j)$$



- 缓解了边界过度平滑问题
- 推断效率低
  - 以像素为结点
  - 结点数和边数巨大

全连接条件随机场

$$E(x)=\sum_i \psi_u(x_i)+\sum_i\sum_{j>i}\psi_p(x_i,x_j)$$

- 高斯边势能

$$\psi_p(x_i,x_j)=\mu(x_i,x_j)\sum_{m=1}^{K}w^{(m)}k^{(m)}(\mathbf{f}_i,\mathbf{f}_j)$$

  - 标记相容性函数 μ
  - 高斯核函数线性组合
  - 任意特征空间

全连接条件随机场

$$\psi_p(x_i,x_j)=\mu(x_i,x_j)\left(w^{(1)}\exp\left(-\frac{|\mathbf{p}_i-\mathbf{p}_j|}{2\theta_\alpha^2}-\frac{|I_i-Ij|}{2\theta_\beta^2}\right)+\right.$$

$$\left.w^{(2)}\exp\left(-\frac{|\mathbf{p}_i-\mathbf{p}_j|}{2\theta_\gamma^2}\right)\right)$$

- 标记相容性
  - $\mu(x_i,x_j)=1_{[x_i\neq x_j]}$
  - $\mu(x_i,x_j)$ 也可以从训练数据中学习
- 外观核函数
  - 颜色
- 局部平滑性
  - 抑制像素级的噪声

全连接条件随机场

□ 推断问题

- ■ 寻找最可能的标记

$$\hat{x} = \operatorname*{argmax}_{x} P(x) \quad \text{其中} \quad P(x) = \exp(-E(x))$$

□ 平均场近似

- ■ 寻找近似分布$Q(x) = \prod_i Q_i(x_i)$使其在KL散度$D(Q||P)$下与$P(x)$接近

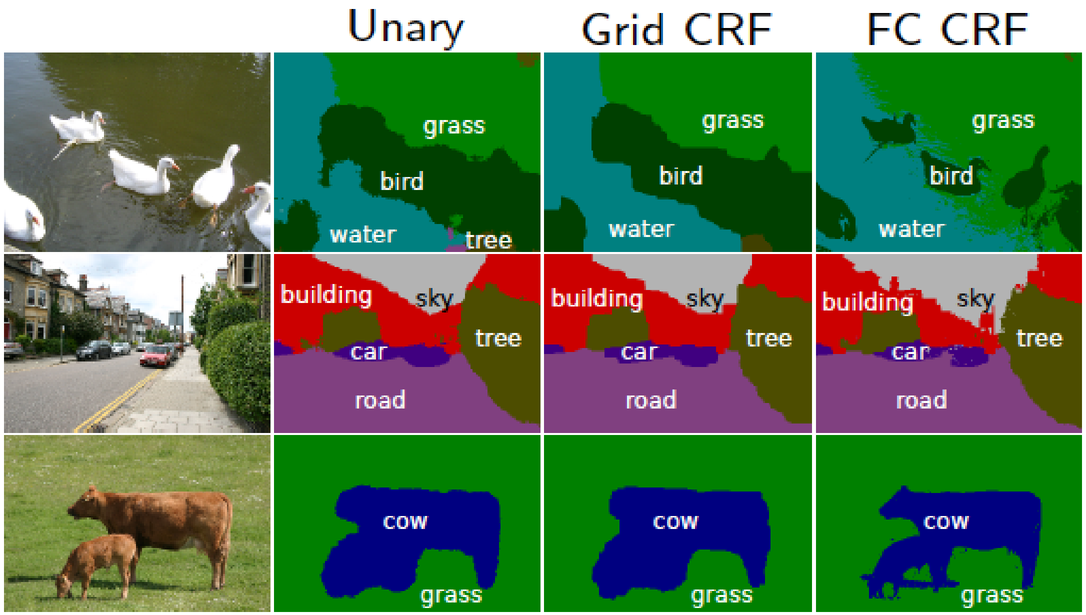
$$Q_i(x_i = l) = \frac{1}{Z_i} \exp\left\{-\psi_u(x_i) - \sum_{l' \in \mathcal{L}} \mu(l, l') \sum_{m=1}^{K} w^{(m)} \sum_{j \neq i} k^{(m)}(\mathbf{f}_i, \mathbf{f}_j)\, Q_j(l')\right\}$$

- ■ $\hat{x}_i \approx \operatorname*{argmax}_{x_i} Q_i(x_i)$

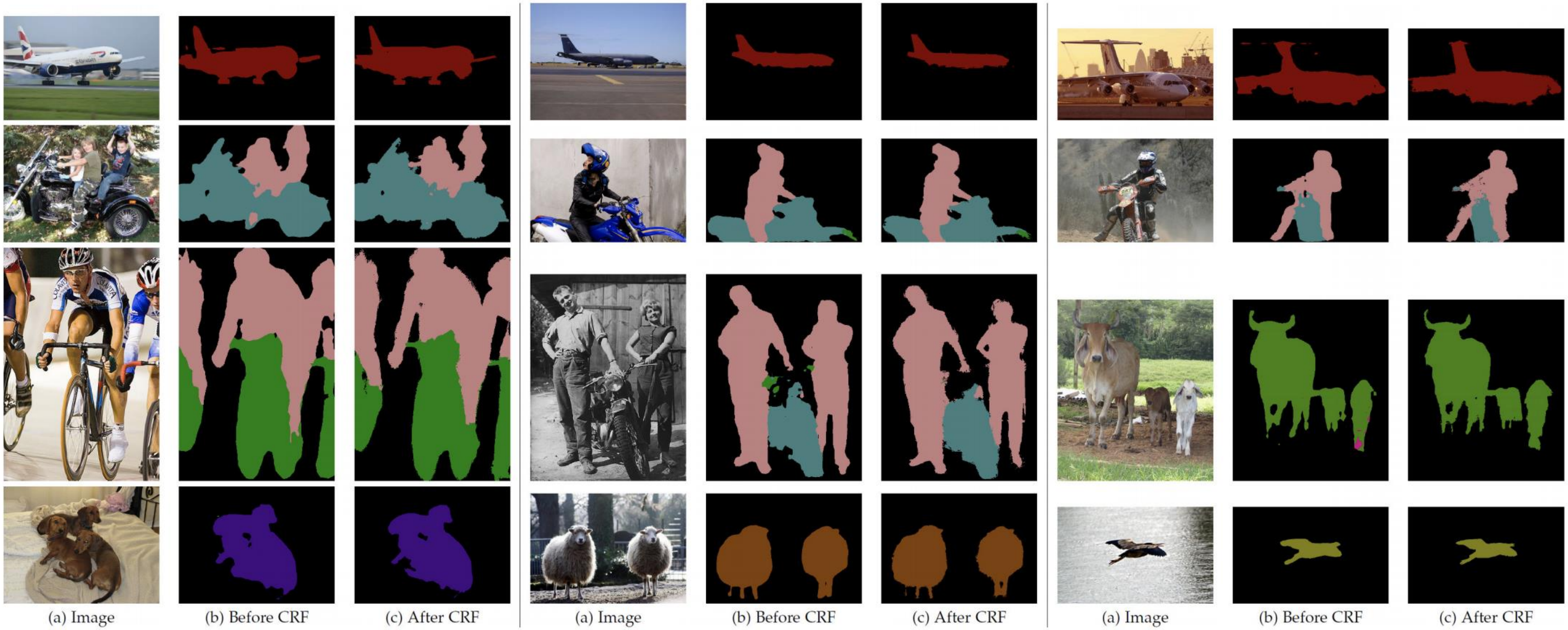
全连接条件随机场

- 实验结果
  - MSRC 数据集

	Time	Global	Avg
Unary	-	84.0	76.6
Grid CRF	1s	84.6	77.2
FC CRF	**0.2s**	**86.0**	**78.3**



全连接条件随机场



Method	before CRF	after CRF
LargeFOV	65.76	69.84
ASPP-S	66.98	69.73
ASPP-L	68.96	71.57

*L.-C. Chen**, G. Papandreou*, I. Kokkinos, K. Murphy, and A. L. Yuille, DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs, PAMI 2018.