在当代科学与社会的广阔天地里，人们都可以看到一种叫作随机过程的数学模型：从银河亮度的起伏到星系空间的物质分布、从分子的布朗运动到原子的蜕变过程，从化学反应动力学到电话通讯理论、从谣言的传播到传染病的流行、从市场预测到密码破译，随机过程理论及其应用几乎无所不在。人类历史上第一个从理论上提出并加以研究的过程模型是马尔科夫链，它是马尔科夫对概率论乃至人类思想发展作出的又一伟大贡献。

一言以蔽之：

马尔科夫过程，是指下一个时间点的指只与当前值有关系，与以前没有关系，即未来决定于现在而不是过去。

用一个通俗的比喻来形容，一只被切除了大脑的白鼠在若干个洞穴间的蹿动就构成一个马尔科夫链。因为这只白鼠已没有了记忆，瞬间而生的念头决定了它从一个洞穴蹿到另一个洞穴；当其所在位置确定时，它下一步蹿往何处与它以往经过的路径无关。这一模型的哲学意义是十分明显的，用前苏联数学家辛钦（1894－1959〕的话来说，就是承认客观世界中有这样一种现象，其未来由现在决定的程度，使得我们关于过去的知识丝毫不影响这种决定性。这种在已知 “现在”的条件下，“未来”与“过去”彼此独立的特性就被称为马尔科夫性，具有这种性质的随机过程就叫做马尔科夫过程，其最原始的模型就是马尔科夫链。

马尔可夫随机场（Markov Random Field）包含两层意思。

马尔可夫性质：它指的是一个随机变量序列按时间先后关系依次排开的时候，第N+1时刻的分布特性，与N时刻以前的随机变量的取值无关。拿天气来打个比方。如果我们假定天气是马尔可夫的，其意思就是我们假设今天的天气仅仅与昨天的天气存在概率上的关联，而与前天及前天以前的天气没有关系。其它如传染病和谣言的传播规律，就是马尔可夫的。

随机场：当给每一个位置中按照某种分布随机赋予相空间的一个值之后，其全体就叫做随机场。我们不妨拿种地来打个比方。其中有两个概念：位置（site），[相空间](http://www.hudong.com/wiki/%E7%9B%B8%E7%A9%BA%E9%97%B4)（phase space）。“位置”好比是一亩亩农田；“相空间”好比是种的各种庄稼。我们可以给不同的地种上不同的庄稼，这就好比给随机场的每个“位置”，赋予相空间里不同的值。所以，俗气点说，随机场就是在哪块地里种什么庄稼的事情。

马尔可夫随机场：拿种地打比方，如果任何一块地里种的庄稼的种类仅仅与它邻近的地里种的庄稼的种类有关，与其它地方的庄稼的种类无关，那么这些地里种的庄稼的集合，就是一个马尔可夫随机场。

Markov Random Fields(MRF)是undirected graph的概率表示，下面说说它在computer vision中的应用。

MRF应用在视觉中，相当于一个Labeling问题，更具体点，是通过MAP inference来确定图中每个节点的label。MRF相比其他方法的优势是：1）提供了一种principled method来对Prior knowledge建模，2）MRF可以很容易用定量的方法描述contextual information。 因此，相比其它pixel-based, 或local-based 方法，它可以考虑到环境知识的影响，如果建立的图模型得当，进而可能获得全局最优解释，这样正是向human vision更靠近了一步。

说到MRF的Inference，首先必须有 graph construction, parameter learning,最后才是Inference，图的创建一般是对问题本身的建模，比如在image restoration 和image segmentation中，常用到4-neighborhood或8-neighborhood的pairwise模型，这样，4-或8-相邻的像素中间便用边连接，这样的模型就是paradigmatic pairwise Markov model,如果要加入高阶（>=3)的potential，相当于我们引入了更多的约束，比如：connectivity 约束、非基督分类结果的约束..., 说到非监督分类结果的约束，要注意的是：一定是其他分类方法，而非MRF本身的分类结果累构成新的约束。

MRF中参数学习方法在此略过，后续补充。

下面重点说MRF inference问题，即解求能量函数最小能量的问题。对于经典的只有unary 和 binary potential的MRF模型，graph cut已经能够在Linear time内进行求解，如果加入更高的potential,虽然问题本身可能变成了NP-hard，仍然有很多近似算法，比如Loopy belief propagation(LBP)，tree-reweighted message passing(TRW),Metropolis-Hastings,MCMC等等。撇开这些方法，其实问题的实质是 energy minimization，值得一提的是CVPR现在有个workshop是专门讨论这个问题的，叫做energy minimization methods in computer vision and pattern recognition (EMMcvpr). 组织者是：Yuri Boykov（UWO）, Fredrik Kahl, Victor Lempitsky ,Vladimir Kolmogorov(UCL), Olga Veksler(UWO),个个都活跃在算法、离散数学、变分法、图论等领域的前沿，所以大家对能量最小化的优化方法感兴趣的可以经常关注这些教授的主页，他们也主要做计算机视觉方面的应用，所以他们的publication会多数发表在ICCV,CVPR,ECCV,NIPS上面。

上面粗要的概述了MRF的三个步骤，下面谈谈自己对MRF的看法： 1）虽然它是通过MAP获得整体的maximal likely solution,但是，如果对问题本身建模时，只考虑到了局部约束，则：MRF得到的结果仍然是局部的，比如考虑最经典的影像分割，如果只用到pairwise Model，则很可能把一个完整的物体分割成了2块甚至更多块。2）当引入更多的constaints时，不同的约束如果非独立，而通常，我们使用Gibbs Distribution来描述Markov Network,不同的potential是相加的关系，转化到概率，则是相乘，就是不独立的两个clique直接相乘了，所以此时存在参数的冗余（redundency),所以Parameter learning时，要解决这个问题 （这个问题在测量数据处理中叫 参数的显著性检测）。