

什么是贝叶斯网络？(What are Bayesian networks?)

甄景贤 (King-Yin Yan)

General.Intelligence@Gmail.com

1 命题和它们之间的关系

上次介绍过「命题逻辑」，**命题**即是能赋予 **真假值** (truth value) 的东西。例如「昨天下雨」、「人是动物」等。

命题之间有「关系 relations」，例如：「吃了不洁食物」 \rightarrow 「肚子痛」；
记作 $P \rightarrow Q$ 或叫「P 蕴涵 Q」。

这个「**箭头**」是一切逻辑中最重要的，有了它才可以有「**推导** deduction」这回事，否则知识变成一堆离散的没关连的点子。

宇宙无限，但人脑有限，以有限的脑怎能理解无限的世界？那是靠 general rules。那些 rules 里面有变量 (variables)，所以逻辑 formulas 可以像 template (模版) 那样套到不同的实例上，然后推出不同的结论，於是「经验世界」可以被**压缩**成有限的资讯。

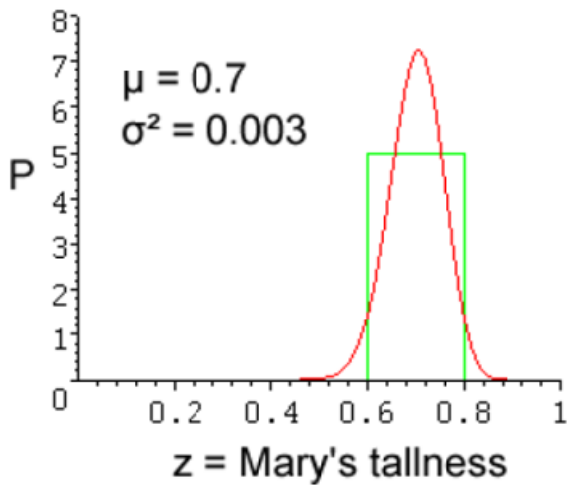
总言之，**变量**是**压缩**的基础，**箭头**是**推导**的基础。

2 机率

但，在经典逻辑中，所有命题都是 **非真即假** (binary) 的，这当然有局限，所以我们这一代的研究者都不用经典逻辑了。我的 Genifer、Ben Goertzel 的 OpenCog、王培的 OpenNARS，我们都设计了某种“uncertainty logic”。



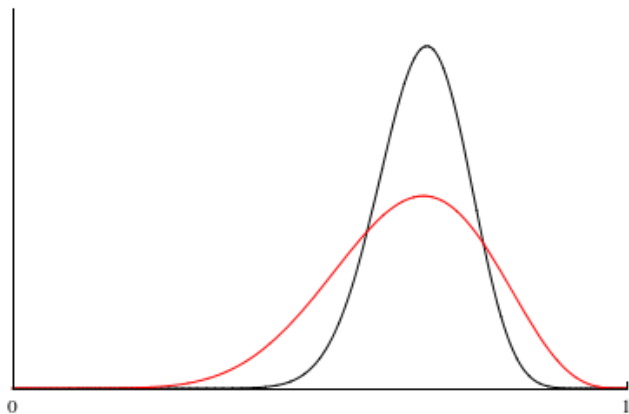
我使用的办法是经典的概率论加上模糊性 (fuzzy)；例如，如果说「玛莉很高」，那概率的分布就可以像下图的红色曲线那样：



(1)

即是说，玛莉的高度，是以不确定的概率分布，而最有可能的高度就是尖端的 peak 位置。注意，我不标示真正的高度，而是用 [0,1] 这区间来表示「高」的程度，例如 0.7 是比平均高的，0.5 则是平均。


但如果不确定性增加的话，那曲线会变得更扁，例如：




(2)






还有，机率特性，就是它分布的曲线下面的总面积一定是 1，因为无论玛莉的高度如何分布，所有可能性的总和就是一个「必然事件」，而必然事件的机率是 1。

温习一下机率是什么？

例如，在桥牌中，大於  的牌，包括：



共 9 张，所以，随意地抽一张而没有作弊的话，大於  的机会就是 9/52。

但如果我们已经知道了 ("given") 那张牌是黑色的，那么可能性的个数就变成了      亦即是 5 张。

在概率论中，这 "given" 的概念很重要。例如医学上可以说：「已知某病人是男性、年龄>40、吸烟 20 年；那他患肺癌的机会是多少」等等。

用符号表示我们的例子：

$$P(> \text{Queen of Hearts}) = 9/52 \approx 17.3\%$$

$$P(> \text{Queen of Hearts} \mid \text{黑色}) = 5/26 \approx 19.23\%$$

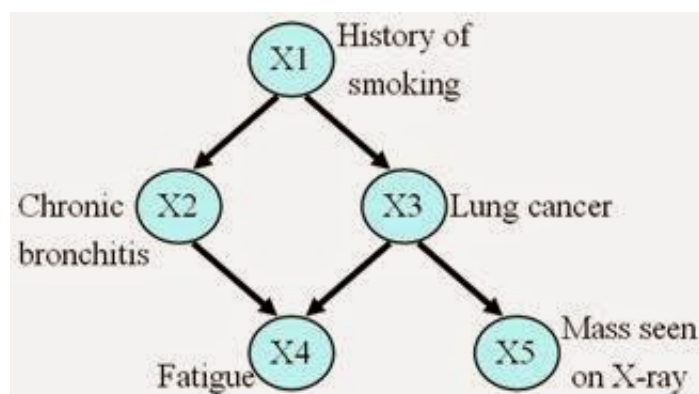
(因为黑色牌有 26 张)

那 $P(A|B)$ 就是 "A given B" 的意思。

(3)

3 Bayesian network

以下是一个 Bayesian network 的例子:



(4)

每一个箭头，如果是由 X 指向 Y，便表示 “Y given X”，亦即是「如果知道了 X 发生的话，Y 的机会是多少」。

例如在图中可看到，吸烟影响肺癌、肺癌影响疲倦、肺癌也影响 X-ray 的结果。

在上图中每个箭头，都附带有一些 $P(X|Y)$ 那样的数据，如下表所示：
(不需深究)

P(X1=no)=0.8	P(X1 = yes)=0.2
P(X2=absent X1=no)=0.95	P(X2=present X1=no)=0.05
P(X2=absent X1=yes)=0.75	P(X2=present X1=yes)=0.25
P(X3=absent X1=no)=0.99995	P(X3=absent X1=no)=0.00005
P(X3=absent X1=yes)=0.997	P(X3=absent X1=yes)=0.003
P(X4=absent X2=absent, X3=absent)=0.95	P(X4=present X2=absent, X3=absent)=0.05
P(X4=absent X2=absent, X3=present)=0.5	P(X4=present X2=absent, X3=present)=0.5
P(X4=absent X2=present, X3=absent)=0.9	P(X4=present X2=present, X3=absent)=0.1
P(X4=absent X2=present, X3=present)=0.25	P(X4=present X2=present, X3=present)=0.75
P(X5=absent X3=absent)=0.98	P(X5=present X3=absent)=0.02
P(X5=absent X3=present)=0.4	P(X5=present X3=present)=0.6

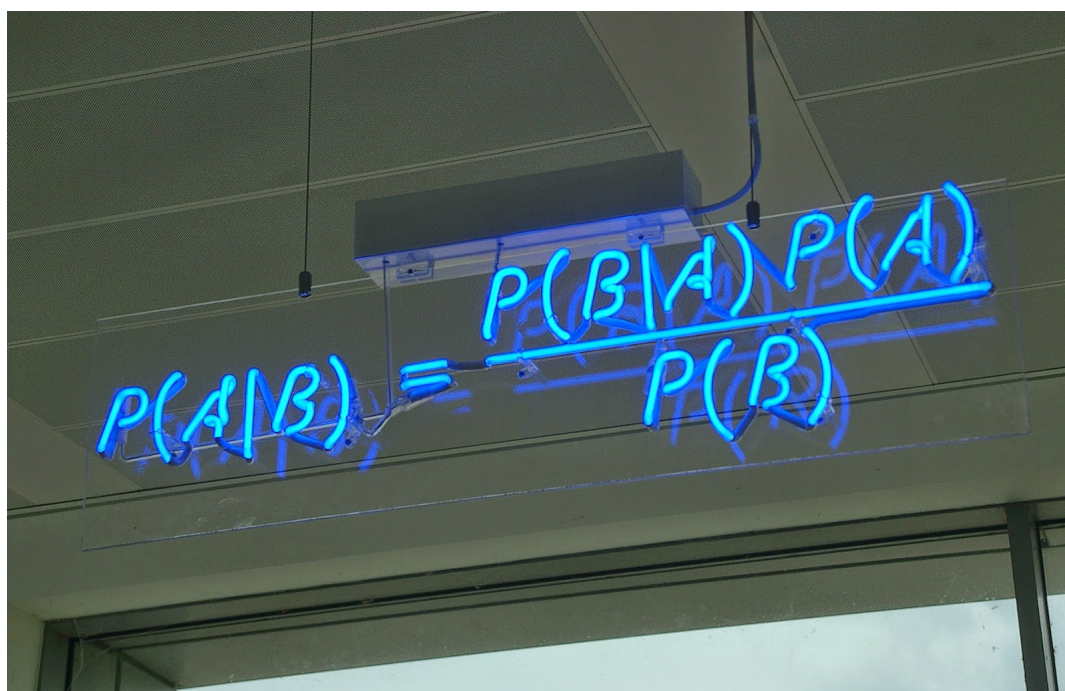
(5)

之所以叫 Bayesian network，是因为 Thomas Bayes 这个和牛顿同期的数学家：



(6)

他发现了一个公式可以用来计算 $P(X|Y)$ 那样的机率:


$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

(7)

亦可以更为对称地写成:

$$P(A|B)P(B) = P(A, B) = P(B|A)P(A) \quad (8)$$

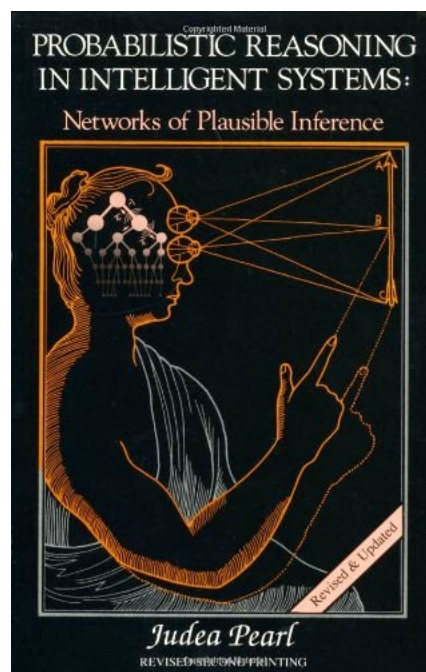
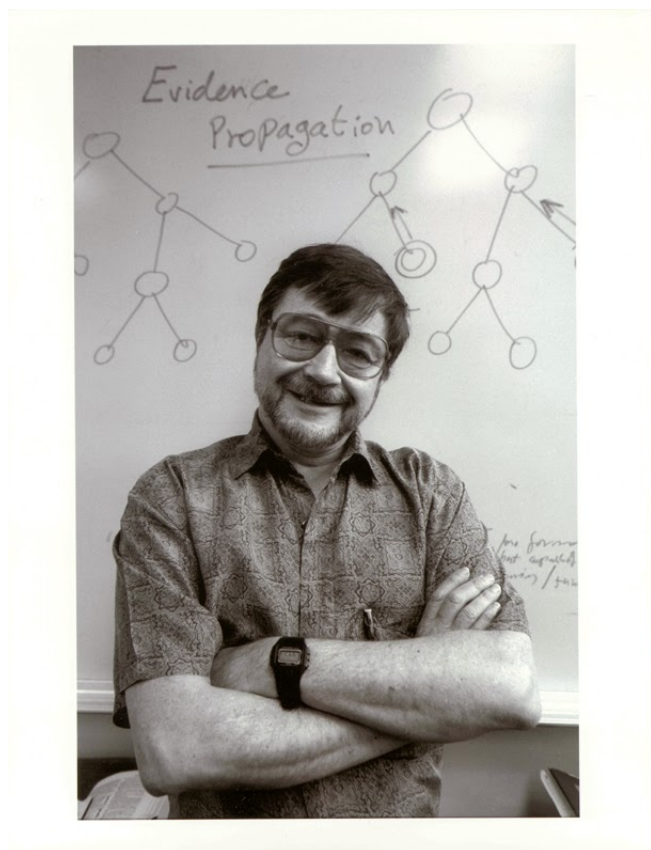
其中 $P(A, B)$ 是”A and B”都发生的机率，和”A given B”不同。

这公式在初中或高中会学到，不是很难明。

贝叶斯网络的难处，是在於网络把不同的机率连系起来了，所以 Bayes rule 要一个套着一个地连续运用，变成很复杂、很多 sum of products。但其实如果有耐性和细心的话，那个算法应该不难，因为除了 Bayes rule 与及机率的基本运算之外，就没有其他了。

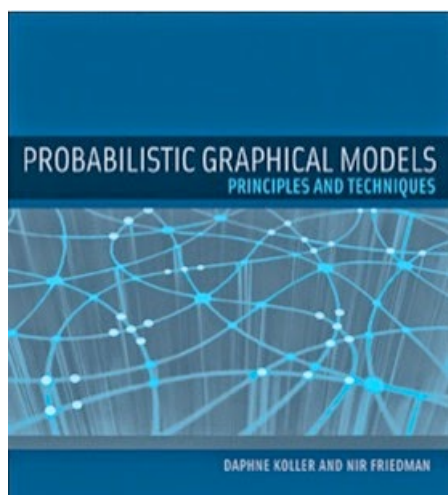
4 参考资料

以前经典 AI 不用机率，直到 1980's Judea Pearl 写了一本很详细分析 Bayesian network 的书，使它变成主流：



Pearl 的儿子是美国驻伊拉克记者，他们家是犹太人，他不幸被恐怖分子捉了而被割首，过程被摄影下来。我不能断定他是极端亲美，还是作为一个同情者而被无辜杀害？

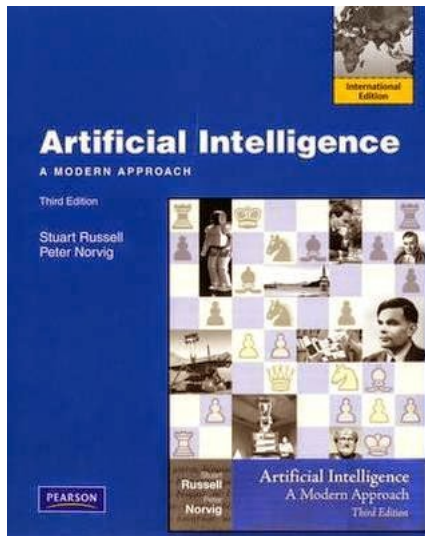
另外这个是 Daphne Koller，她是这本新的巨著 (2009 年，1280 pages!) 的作者之一：



但这本书太 technical 和详细，对初学者不好读。

在 Stanford U 的 Daphne 是 Coursera 的创始人之一，她的 (免费的) 课 <https://www.coursera.org/course/pgm> 详细地讲解 Bayesian network 的算法和学习法。

顺带一提，这个是 Peter Norvig，他和 Stewart Russell 合著的”AIMA” 是现时最好的人工智能教科书。



— 完 —