电子科技大学
University of Electronic Science and Technology of China

Chap. 3-4
图模型的两种表示方法

黄峰 & 黄晨

Data Mining Lab,
Big Data Research Center, UESTC
Email：huangchen.uestc@gmail.com

目录

数据挖掘实验室
Data Mining Lab

- **第一部分：贝叶斯网络**
  - ✓ 啦啦
  - ✓ 啦啦啦
  - ✓ 啦啦啦
  - ✓ 。。。
- **第二部分：马尔科夫随机场**
  - ✓ MRF的初步认识
  - ✓ 越来越熟悉的MRF
  - ✓ 熟悉到陌生的BN和MRF

友情提示：为了让大家更好的理解，我们选择用中文做ppt，并且首次加入新的风格，避免枯燥的理论讲解，迈出人类坚实的一小步。

第一部分

贝叶斯网络

装逼之路

傻逼明明的故事

假设我们有5个随机变量，分别是Difficulty(D), Intelligence(I), SAT(S), Grade(G), Letter(L)变量。

其中的故事是这个样子的： 明明考啦邵老师的数据挖掘课的考试，成绩是G，这门考试的难度是D，然后他的托福成绩是S，傻逼明明拿着成绩去求杨老师给他写推荐信（L）？(其中G的取值a,b,c，其余皆是二元变量)

问题： D, I, S, G, L的联合概率分布的参数有多少个？

没错， 就是$2^4 \times 3-1$

装逼之路

数据挖掘实验室
Data Mining Lab
DM LESS IS MORE

肿么可以这么low呢，我们要装逼！

$$P(D,I,S,G,L) = P(D)P(I|D)P(S|D,I)P(G|D,I,S)P(L|D,I,S,G)$$

这样的话，我们求解参数会不会发生变化呢？

$P(D)$ 1个
$P(I|D)$ 2个
$P(S|D,I)$ 4个
$P(G|D,I,S)$ $2*8=16$个
$P(L|D,I,S,G)$ 24个

劳资辣么萌，肿么可能会装错逼 ◜(◞ ^ ◟)◝

$1+2+4+16+24 = 47 = 2^4 \times 3-1$=装逼失败？

装逼之路

你知道吗，一个大问题变成啦几个小问题

难道这逼还不算成功？不要只见参数数量木有变，而不见将来之扩展。万一傻逼明明后续问，我拿到推荐信之后，傻逼大学有多大的可能性要我，那不就萌逼啦。

模块化的思想
扩展性啊！！！

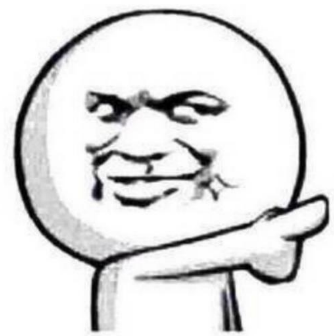
装逼之路

当然这个逼装的还不够，让我们好好回忆下刚刚的故事。

傻逼明明很有心机地只给了杨老师考试成绩，所以杨老师只能根据这个判断给不给推荐，所以和他是不是傻逼，考试难不难，SAT的成绩好像半毛钱关系都木有。

也就是P(L|D,I,S,G) = P（L|G），也就是说这里我们只要3个参数就行啦，装逼的机会来啦。

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01



后退，我要开始装逼了

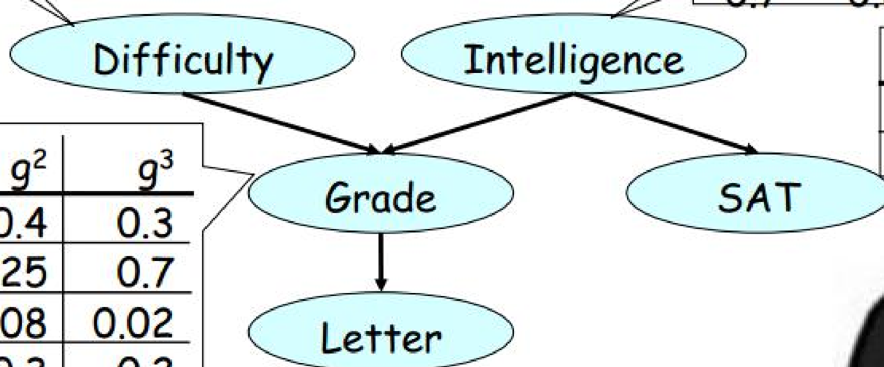
如果说我了解一些其他的独立关系：

$$P(D,I,S,G,L) = P(D)P(I|D)P(S|D,I)P(G|D,I,S)P(L|D,I,S,G)$$

$$P(D,I,S,G,L) = P(D)P(I)P(S|I)P(G|D,I)P(L|G)$$

d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3



	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

今天这逼，
我装定啦。

如何系统的装逼

装逼的语言：

晨晨说过：“某某人说过，图模型是一种语言，刻画变量之间的关系”

那么贝叶斯网络语言的语义是什么呢？

对于每个变量X_i（$X_i \perp \text{NonDescendants}_{X_i} \mid \text{Pa}^{g}_{X_i}$）

就是在已知X_i的父节点的情况下，X_i与不是X_i后代的节点是独立的。

如何系统的装逼

装逼实质：

贝叶斯网络的核心在于用一个有向无环图对变量之间的依赖关系进行编码。

人话：把一个求联合概率分布的问题根据各个变量之间的独立关系变成了很多个小的条件概率分布和边缘概率分布的问题。

如何系统的装逼

装逼的形式化：从现象到抽象

$$P(D,I,S,G,L) = P(D)P(I)P(S|I)P(G|D,I)P(L|G)$$

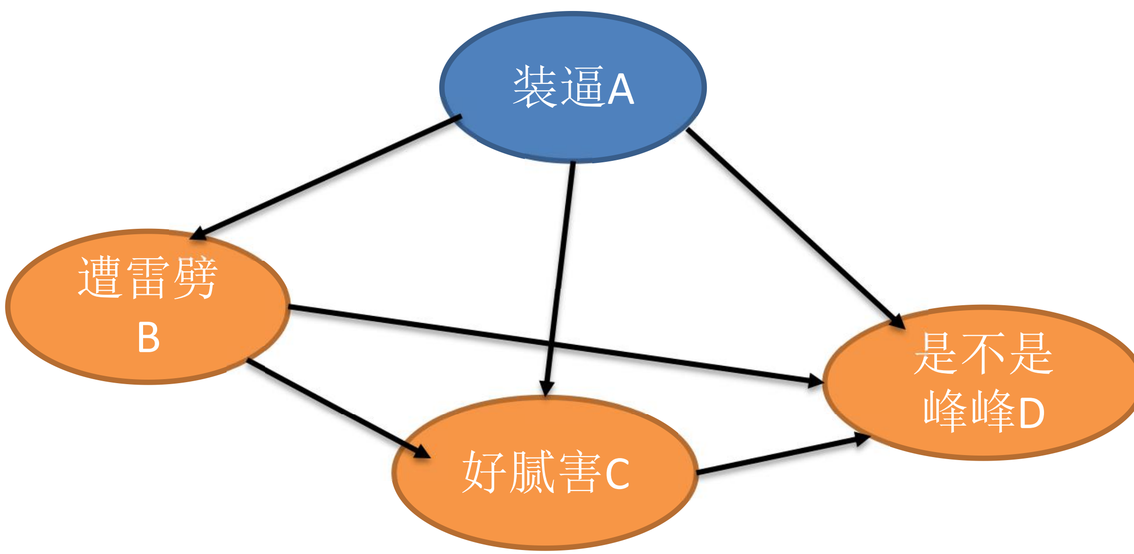
贝叶斯网络的一般链式规则：

$$P(X) = \prod_{i=1:d} P(X_i|Parents(X_i))$$

等我们把一些B捡起来，在装。

如何判断是不是装逼（插叙）

有个学啦数据挖掘的人，叫明明，他想预测一个人是不是装逼，于是他就分析出几个因素，1,那个人是不是遭雷劈啦，2，那个人是不是很厉害，3，那个人是不是峰峰。

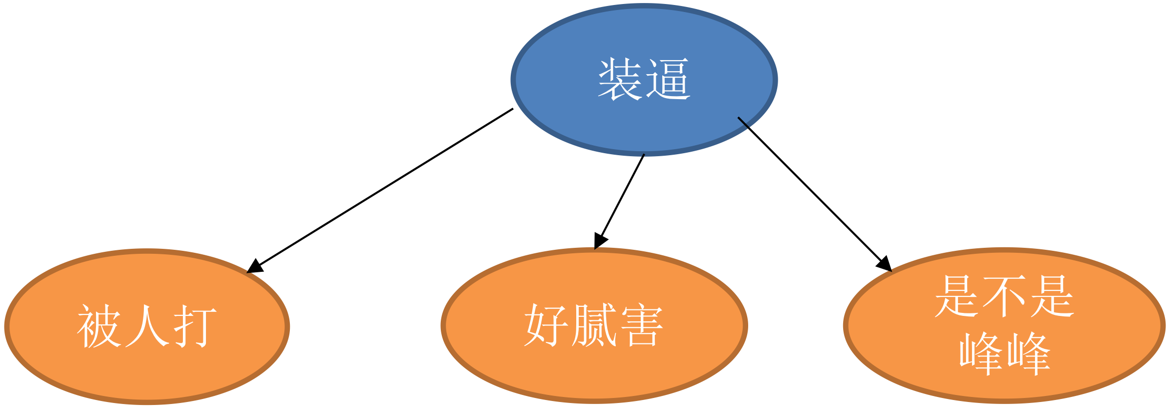


请允许我
卖个萌

朴素的装逼 Naïve bayes

朴素贝叶斯是一种特殊的贝叶斯网络！

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

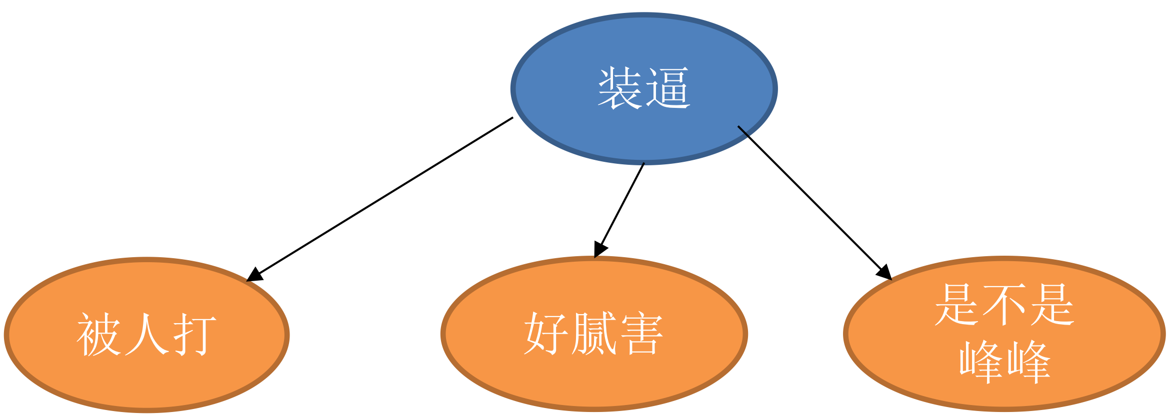


朴素的装逼 Naïve bayes

朴素贝叶斯是一种特殊的贝叶斯网络！

问题1：如果求的不是 装逼？

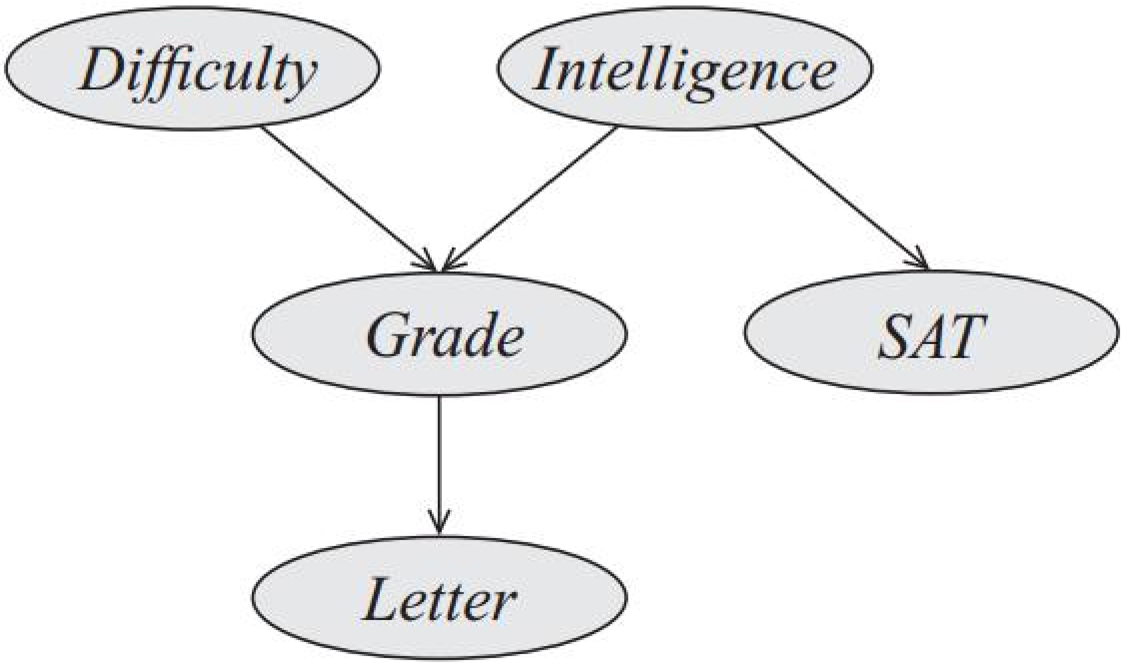
问题2：如果有缺失值？ $\frac{P(A)P(C|A)P(B|A)}{P(B,C)}$



装逼在继续 Bayes network

数据挖掘实验室 Data Mining Lab

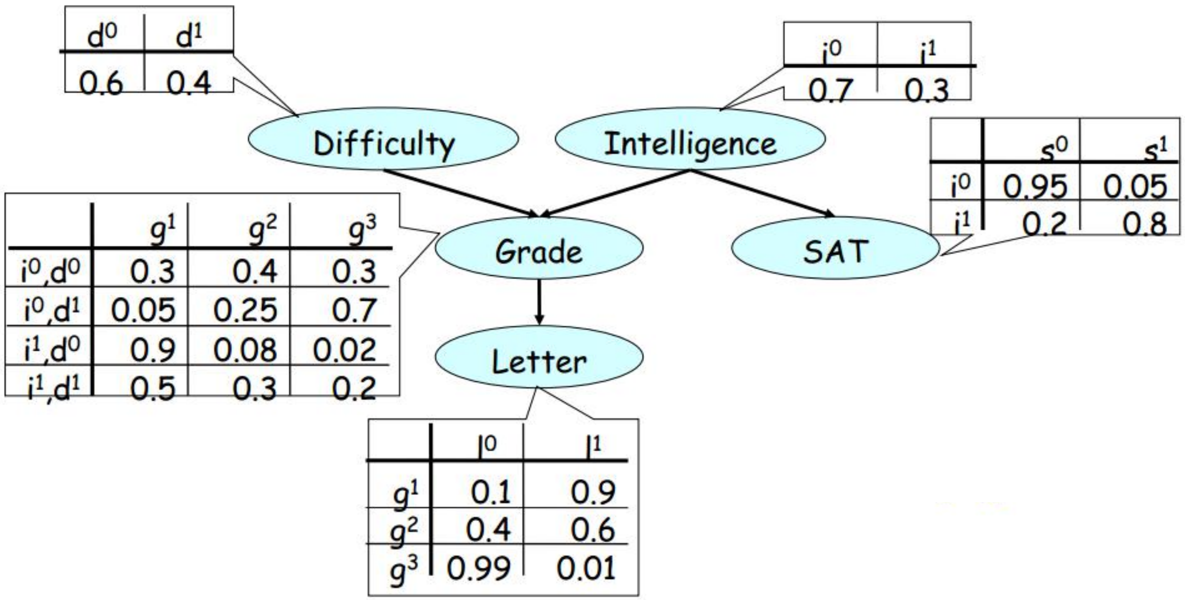
傻逼明明，我们又回来啦！！！！



已知杨老师心地善良，给啦傻逼明明一发棒棒哒的推荐信，那么请问傻逼明明是傻逼的概率是多少呢？

装逼在继续

贝叶斯网络上的推断：



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

$$P(i|l)=\frac{P(i,l)}{P(l)}=\frac{\sum_{D,I,S}P(D,G,S|L=l,I=i)}{\sum_{D,I,G,S}P(D,I,G,S|L=l)}$$

装逼新问题

数据挖掘实验室
Data Mining Lab

为什么我要重复计算那么多无用的东西呢？

依赖性和独立性是分布的两个主要性质，对理解分布的特征非常重要。正如即将看到的，独立性质在回答查询时非常重要。它们可以被用于降低推断的计算成本。

疯言疯语：
如果我们能把推断的时候的变量影响范围根据独立性确定，那么是不是可以在一个变量子集里做推断？

这个逼装的眼前一亮！！！

装逼新概念（1） Active Trails

数据挖掘实验室
Data Mining Lab

因果迹

- X → Z → Y

证据迹

- X ← Z ← Y

共同的原因

- X ← Z →Y

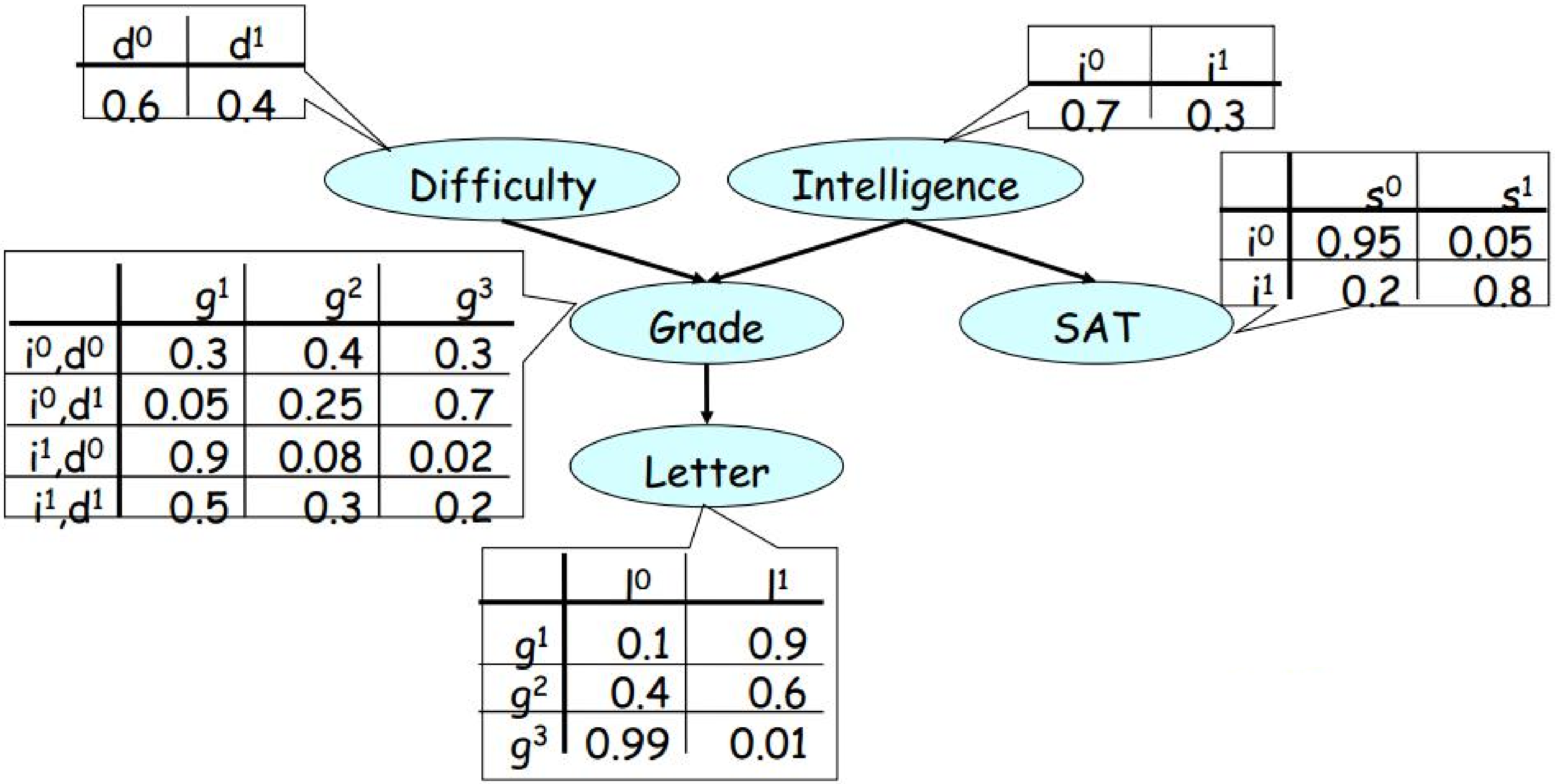
当z没有被观察到时有效

共同的作用

- X → Z ← Y

当且仅当观察到z或z的后代的时候有效。

装逼新概念（1）Active Trails



d^0	d^1
0.6	0.4

i^0	i^1
0.7	0.3

	g^1	g^2	g^3
i^0,d^0	0.3	0.4	0.3
i^0,d^1	0.05	0.25	0.7
i^1,d^0	0.9	0.08	0.02
i^1,d^1	0.5	0.3	0.2

	s^0	s^1
i^0	0.95	0.05
i^1	0.2	0.8

	l^0	l^1
g^1	0.1	0.9
g^2	0.4	0.6
g^3	0.99	0.01

装逼新概念（2）D-separation

令X,Y,Z是图G中的三个节点集，在给定Z的条件下，假如在X，Y中的任意节点($x \in X, y \in Y$)不存在有效迹，那么X和Y在Z的条件下是d-分离的，x和y是独立的。

$$I(G) = \{x \perp y | Z : d - sep(X; Y | Z)\}$$

装逼的极限在哪里?

贝叶斯网络表述能力的限制

有向图网络的限制性，对独立关系的表达能力不足!

如同线性模型的语言是线，如果数据不线性可分，它就废啦。每一种语言都有自己的表达上的限制。

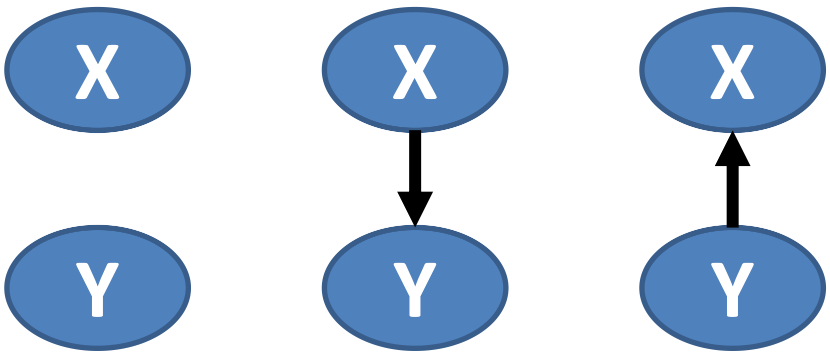
装逼新概念（3）I-Map

对一个多变量的分布P，以及它的一个贝叶斯网络结构图：

- 图G中蕴含的独立关系集合是分布P的独立关系集合的子集，那么G是I（P)的一个I-Map。
- 如果两个结构图中包含的独立关系是相等的，那么他们就是I等价。

装逼新概念（3）I-Map

P中的一些依赖关系不能被G表达：



$I(G) = \{X \perp Y\}$　　$I(G) = \varnothing$　　$I(G) = \varnothing$

X	Y	$P(X,Y)$
x^0	y^0	0.08
x^0	y^1	0.32
x^1	y^0	0.12
x^1	y^1	0.48

X	Y	$P(X,Y)$
x^0	y^0	0.4
x^0	y^1	0.3
x^1	y^0	0.2
x^1	y^1	0.1

装逼新概念（4）潜在分布

G是定义的一个贝叶斯网络，假设分布P能够表示成

$$P(X_1,\cdots,X_n)=\prod_{i=1}^{n}P(X_i\mid \mathrm{Pa}_{X_i}^{\mathcal{G}})$$

那么P是G的潜在分布, 也就是G在同一空间上的因子分解。

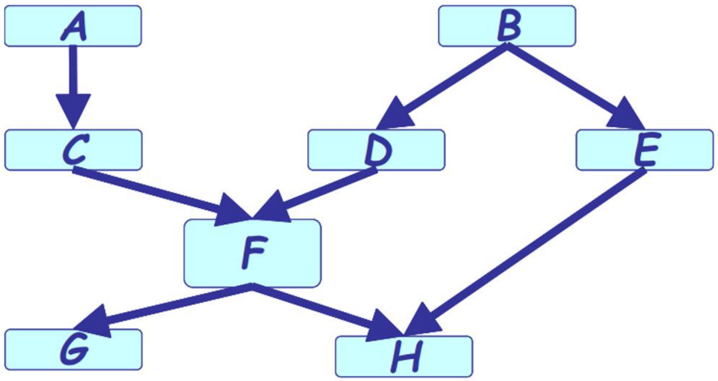
装逼绕口令：

由图的独立性原则（核心语义）能够得到因子分解，同时因子分解也能推得图结构里面的独立性原则，基本独立性又能推出D独立性，所以潜在分布能够包含图G的所有独立性，所以G是潜在分布的Imap。

装逼需要证明吗

数据挖掘实验室
Data Mining Lab

一个示例性质的小说明



Target: $P(H|EF) = P(H|ABCDEFG)$

$$P(H|ABCDEFG) = \frac{P(A\ to\ H)}{P(A\ to\ G)} = \frac{P(A\ to\ H)}{\sum_H P(A\ to\ H)}$$

$$= \frac{P(A)P(B)P(C|A)P(D|B)P(E|B)\ P(F|CD)P(G|F)P(H|EF)}{\sum_H P(A)P(B)P(C|A)P(D|B)P(E|B)\ P(F|CD)P(G|F)P(H|EF)}$$

$$= \frac{P(A)P(B)P(C|A)P(D|B)P(E|B)\ P(F|CD)P(G|F)P(H|EF)}{P(A)P(B)P(C|A)P(D|B)P(E|B)\ P(F|CD)P(G|F)\sum_H P(H|EF)}$$

$$= P(H|EF)$$

装逼终极概念 可靠性完备性

数据挖掘实验室
Data Mining Lab

完备性：

D分离得出的独立性，潜在分布P都满足。

可靠性：

D分离木有得出的X，Y的独立性，潜在分布中P中X，Y相互就依赖。

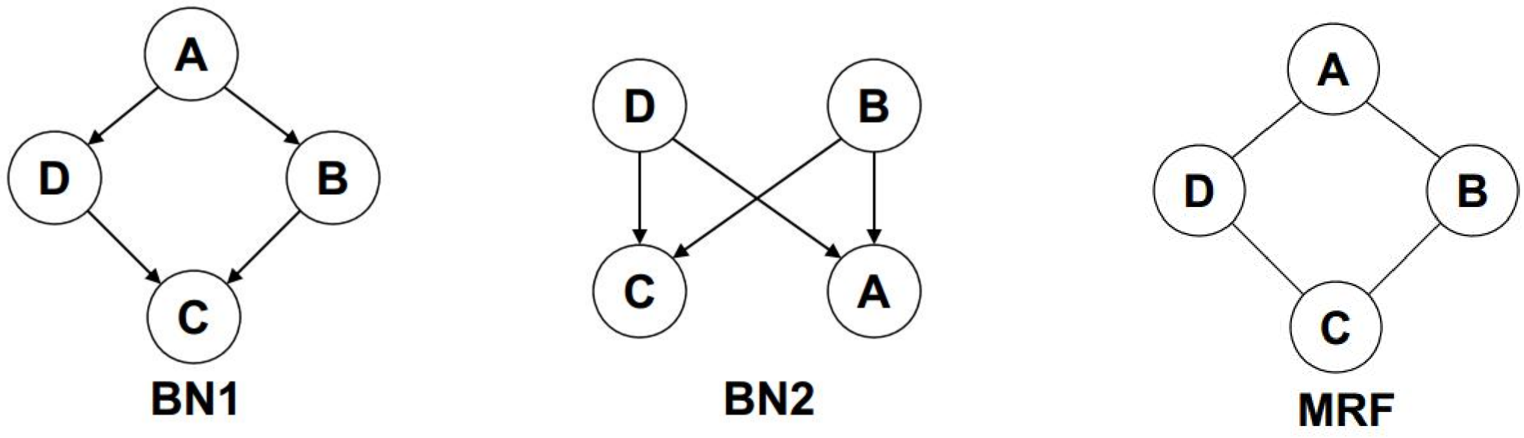
装逼新概念（5）P-Map

如果图G和真实分布P都描述了一样的非独立关系，也就是G和P是I等价的，那么G是P的P-Map.

如果我们找不到P的Map，就是说这个分布不能被贝叶斯网络所完美刻画。

装逼的极限在哪里

假设我们有个分布P，个变量A,B,C,D,它们满足的关系 $(A \perp C) \mid \{B,D\}$ and $(B \perp D) \mid \{A,C\}$，能有个贝叶斯网络画出来吗？



谢谢

这个逼我给自己103分，
多一分宽容，多一分感动，
多一分伙伴们对我的爱。

电子科技大学
University of Electronic Science and Technology of China

多谢

By 黄峰 & 黄晨

Data Mining Lab,
Big Data Research Center, UESTC
Email：huangchen.uestc@gmail.com