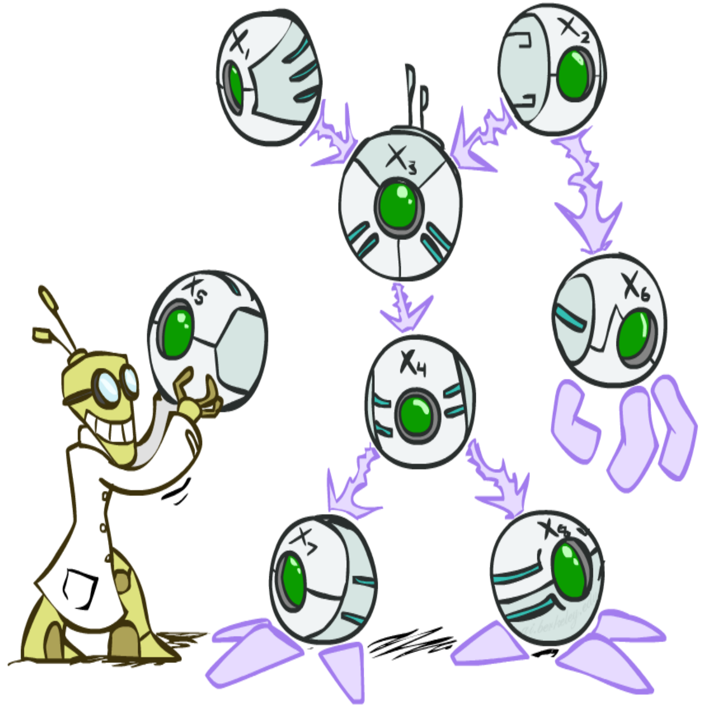
贝叶斯网络(BAYES NETS)



接下来的内容

- 独立性(Independence)
- 条件独立性(Conditional independence)
- 贝叶斯网络(Bayes nets)
  - 语法和语义

概率基础总结，继续

- 求和消除：$P(X=x) = \sum_y P(X=x, Y=y)$
- 条件概率分布：$P(X|Y) = P(X,Y)/P(Y)$

通过列举法概率推理：$P(Q|e_1,...,e_k) =$

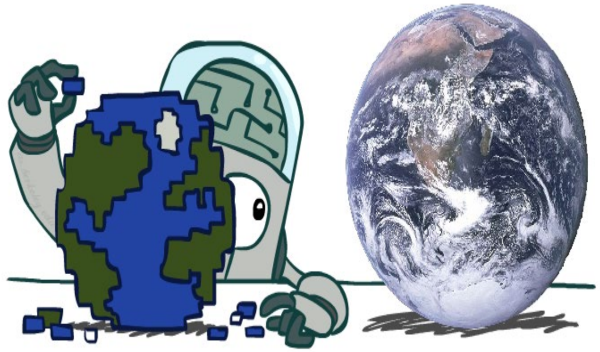
$\alpha \sum_{h_1,...,h_m} P(Q, e_1,...,e_k, h_1,...,h_m)$

- 这里 α（即是 $1/Z$）是一个正规化因子，使得 $P(Q|\cdots)$ 之和为 1

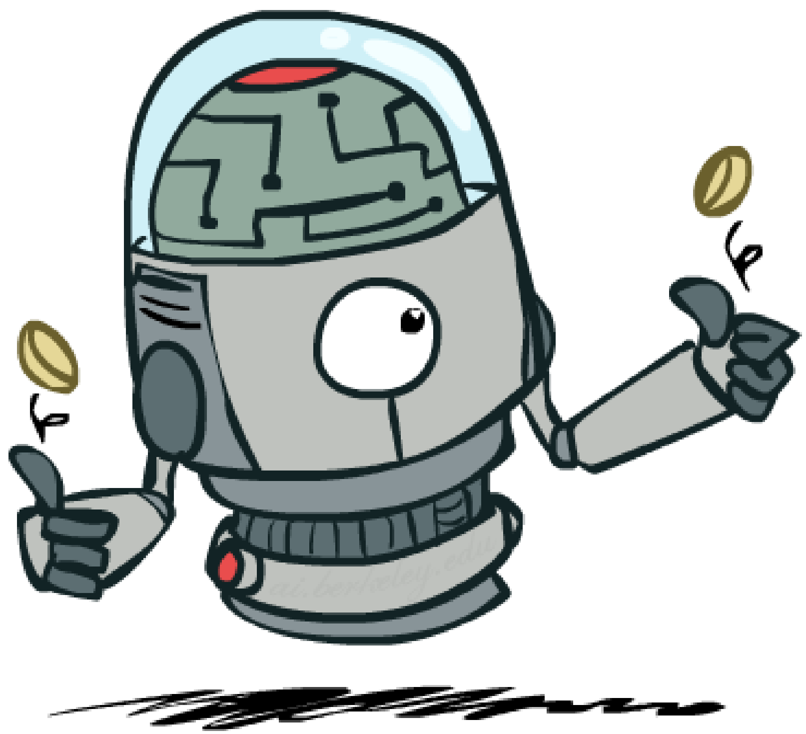
- 乘法规则: $P(X|Y)P(Y) = P(X,Y) = P(Y|X)P(X)$
  - 推广到连锁法则: $P(X_1,..,X_n) = \prod_i P(X_i | X_1,..,X_{i-1})$
- 贝叶斯规则: $P(X|Y) = P(Y|X)P(X) / P(Y)$

概率模型(Probabilistic Models)

- 模型描述的是世界（或某一部分）是如何工作的
- 模型总是一种简化
  - 可能忽略了某些变量和它们之间的交互关系
  - "所有模型都是错的; 但某些是有用的."
    – George E. P. Box
- 概率模型能用来做什么?
  - 我们(或我们的人工智能体) 需要对未知变量进行推理，当给定一些证据后
  - 例如: 解释 (诊断推理)
  - 例如: 预测 (因果推理)
  - 例如: 基于期望利益值的决策
- 如何建立模型，并避免 d^n 的复杂性?



独立性

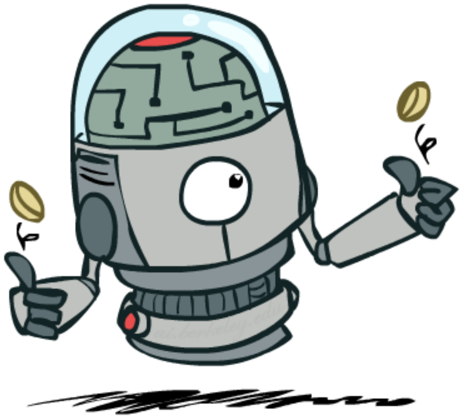


随机变量的独立性

- 两个随机变量是相互独立的 if:
  - 他们的联合概率分布 *分解为* 两个相对简单的分布的乘积
  - 另一种表达形式: $\forall x, y : P(x, y) = P(x)P(y)$

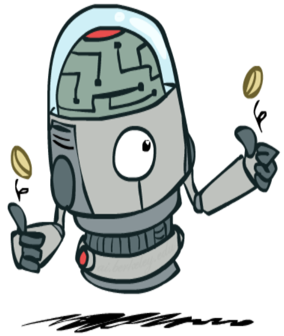
$$\forall x, y : P(x|y) = P(x)$$

  - 记为: $X \perp\!\!\!\perp Y$



- 独立性是构建模型时常用的假设，可以使模型相对简单
  - *Empirical* joint distributions: at best “近似于” 独立的（关联性很小可以忽略）
  - 我们可以对这4个变量构成的模型有什么样的假设: {Weather, Traffic, Cavity, Toothache}?

独立性



- 两个变量 X 和 Y 是 *独立的* 如果

$$\forall x,y \qquad P(x,y) = P(x)P(y)$$

  - 这说明他们的联合分布 *因式分解* 两个简单的分布之乘积

  - 结合乘法规则： $P(x,y) = P(x|y)P(y)$ 我们可以获得另一种形式:

$$\forall x,y\ P(x|y) = P(x) \quad \text{or} \quad \forall x,y\ P(y|x) = P(y)$$

- 举例: 两个骰子 $Roll_1$ 和 $Roll_2$
  - $P(Roll_1=5, Roll_2=5) = P(Roll_1=5)P(Roll_2=5) = 1/6 \times 1/6 = 1/36$
  - $P(Roll_2=5 \mid Roll_1=5) = P(Roll_2=5)$

天气和温度（的独立性）？

计算边缘分布

验证独立性，与P1对比

$P(T)$

T	P
hot	0.5
cold	0.5

$P_1(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$P_2(T,W)$

T	W	P
hot	sun	0.3
hot	rain	0.2
cold	sun	0.3
cold	rain	0.2

$P(W)$

W	P
sun	0.6
rain	0.4

举例: 独立性

n 个公平，独立的硬币翻转:

$P(X_1)$

H	0.5
T	0.5

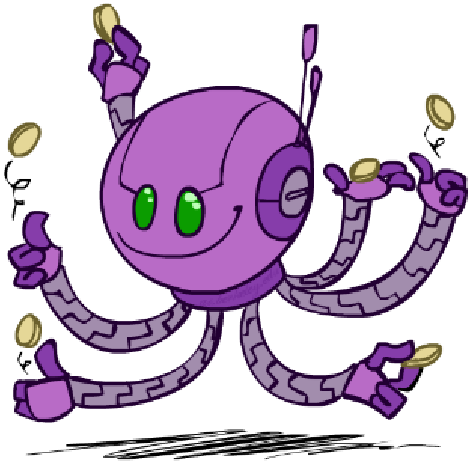
$P(X_2)$

H	0.5
T	0.5

. . .

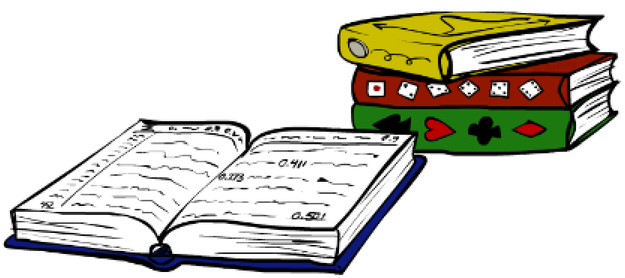
$P(X_n)$

H	0.5
T	0.5



$P(X_1,X_2,\ldots,X_n)$

2^n



条件独立性

- 无条件的 (绝对的) 独立性 是非常少的 (why?)
- *条件独立性* 是关于不确定环境的最基本和相对可靠的建模假设.
- X 条件独立于 Y 当给定 Z

$$X \perp\!\!\!\perp Y \mid Z$$

当且仅当:

$$\forall x, y, z : P(x, y \mid z) = P(x \mid z) P(y \mid z)$$

等价地, 当且仅当

$$\forall x, y, z : P(x \mid z, y) = P(x \mid z)$$

举例：幽灵探测者

- 变量和值域:
  - G (幽灵位置) 在 {(1,1),…,(3,3)}
  - $C_{x,y}$ (在方格 x,y 探测到的颜色；颜色越深离幽灵越近)： {红red,橙orange,黄yellow,绿green}

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

- 假设我们有两个概率分布:
  - ***先验概率 (Prior distribution)*** 幽灵的位置: $P(G)$
    - 假设是均匀(uniform)分布
  - ***传感器模型 (Sensor model)***: $P(C_{x,y} \mid G)$ (只依赖于到G的距离)
    - 例如 $P(C_{1,1} = \text{yellow} \mid G = (1,1)\,) = 0.1$

幽灵搜寻者 计算: 第一步

当检测到 $C_{1,1}$ = yellow 时， 幽灵可能在哪?

换句话说，概率分布 $P(G \mid C_{1,1} = \text{yellow})$?

正规化
Normalize

可能性Likelihood

先验
Prior

Bayes’ rule:

- $P(G \mid C_{1,1} = \text{yellow}) = P(C_{1,1} = \text{yellow} \mid G)\, P(G) \,/\, P(C_{1,1} = \text{yellow})$
  - $= \alpha\, P(C_{1,1} = \text{yellow} \mid G)\, P(G)$ (贝叶斯***Bayesian 更新***)

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
0.01	0.09	0.17

幽灵搜寻者 计算: 第二步

- 当看到 $C_{1,1}$ = yellow, $C_{3,1}$ = green， 幽灵在哪?
- 换句话说， $P(G \mid C_{1,1}$ = yellow, $C_{3,1}$ = green) 概率是什么？
- 我们的模型给出了 $P(C_{x,y} \mid G)$ 需要应用贝叶斯规则:
  - $P(G \mid C_{1,1}=y, C_{3,1}=g)$

    $=\alpha\, P(C_{1,1}=y, C_{3,1}=g \mid G)\, P(G)$

    $= \alpha\, P(C_{3,1}=g \mid C_{1,1}=y, G)\, P(C_{1,1}=y \mid G)\, P(G)$

应用乘法规则 | G

  - 条件后多了一项

0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
0.01	0.09	0.17

?	?	?
?	?	?
?	?	?

幽灵搜寻者 计算: 第二步

- 如何计算 $P(C_{3,1}=g \mid C_{1,1}=y, G)$?
- ***给定幽灵的位置***, 在位置 1,1 观察到的黄色是否影响到在 3,1 为绿色的概率?
  - 不影响!
  - (只依赖于到幽灵的距离)
- 当给定幽灵的位置后，在位置 3,1的颜色是 条件独立（无关的）(***conditionally independent***) 对于在位置1,1的颜色
- $P(C_{3,1}=g \mid C_{1,1}=y, G) \quad = \quad P(C_{3,1}=g \mid G)$

幽灵搜寻者 计算: 第二步

- 观察到 $C_{1,1}$ = yellow, $C_{3,1}$ = green 后，幽灵的位置在哪?
- $P(G \mid C_{1,1} = \text{yellow}, C_{3,1} = \text{green})$?
- 我们的模型给出 $P(C_{x,y} \mid G)$ 只需应用 Bayes’ rule:
  - $P(G \mid C_{1,1}=y, C_{3,1}=g)$

$$= \alpha\, P(G)\, P(C_{1,1}=y, C_{3,1}=g \mid G)$$
$$= \alpha\, P(G)\, P(C_{1,1}=y \mid G)\, P(C_{3,1}=g \mid C_{1,1}=y, G)$$
$$= \alpha\, P(G)\, P(C_{1,1}=y \mid G)\, P(C_{3,1}=g \mid G)$$

依据 $C_{3,1}$ 和 $C_{1,1}$条件独立（无关性），当给定 G后

距离	P(green | 距离)
0	0.01
1	0.1
2	0.2
3	0.3
4	0.4

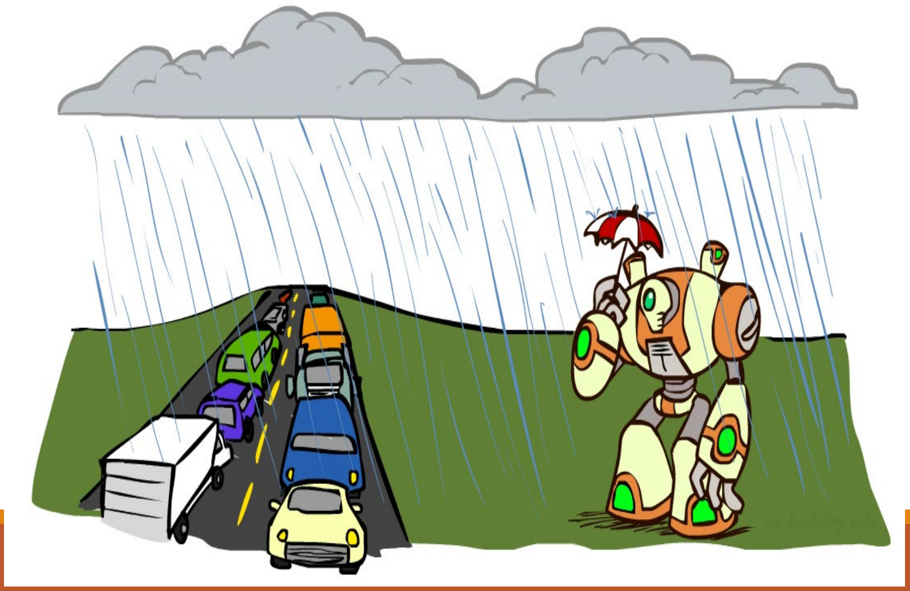
0.11	0.11	0.11
0.11	0.11	0.11
0.11	0.11	0.11

0.17	0.10	0.10
0.09	0.17	0.10
0.01	0.09	0.17

0.34	0.15	0.10
0.13	0.17	0.05
0.01	0.04	0.01

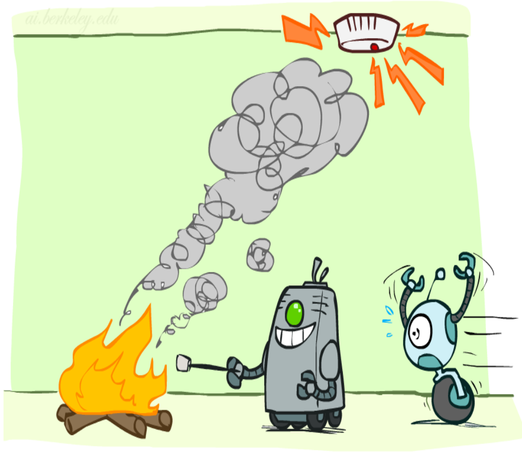
条件独立性（举例）

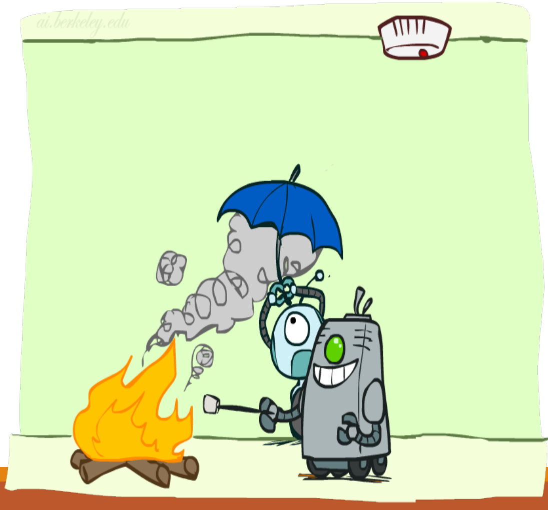
- 关于以下环境的独立性是怎样的:
  - 交通流量
  - 雨伞
  - 下雨



条件独立性

- 关于以下环境的独立性是怎样的:
  - 燃火
  - 冒烟
  - 报警器





条件独立性与连锁法则(Chain Rule)

- 连锁法则

$$P(x_1, x_2,\ldots, x_n) = \prod_i P(x_i \mid x_1,\ldots, x_{i-1})$$

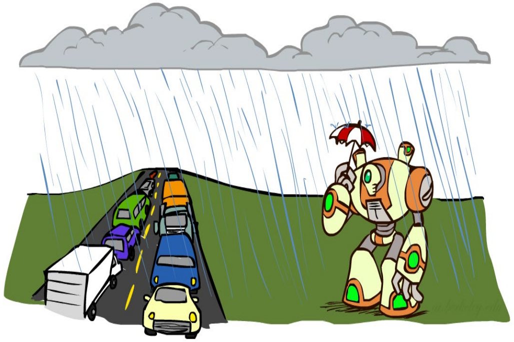
- 简单的分解:

$P(Rain, Traffic, Umbrella) = P(Rain)\ P(Traffic \mid Rain)\ P(Umbrella \mid Rain, Traffic)$

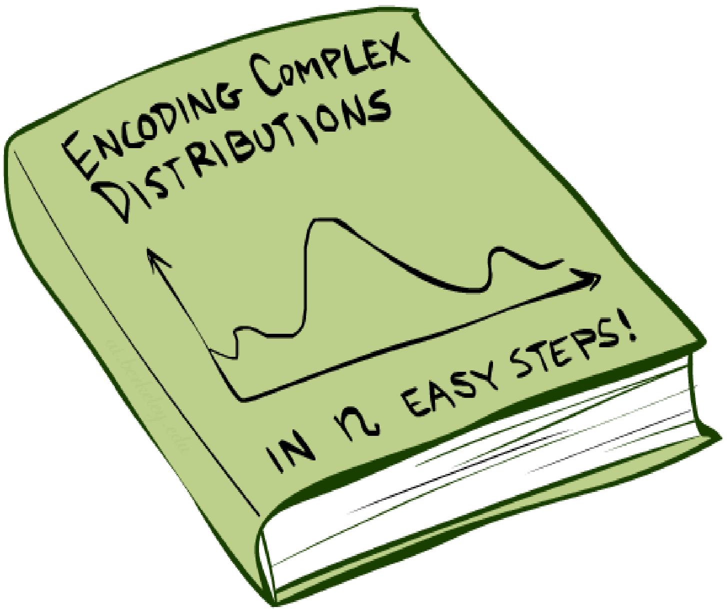
- 利用了条件独立性的假设后:

$P(Rain, Traffic, Umbrella) = P(Rain)\ P(Traffic \mid Rain)\ P(Umbrella \mid Rain)$

- 贝叶斯网络 / 图形模型 帮助表达条件独立性的假设



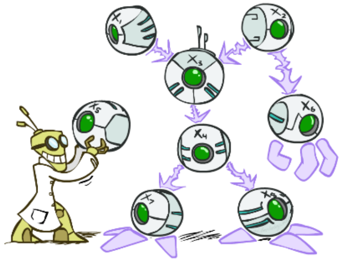
贝叶斯网络(Bayes Nets)



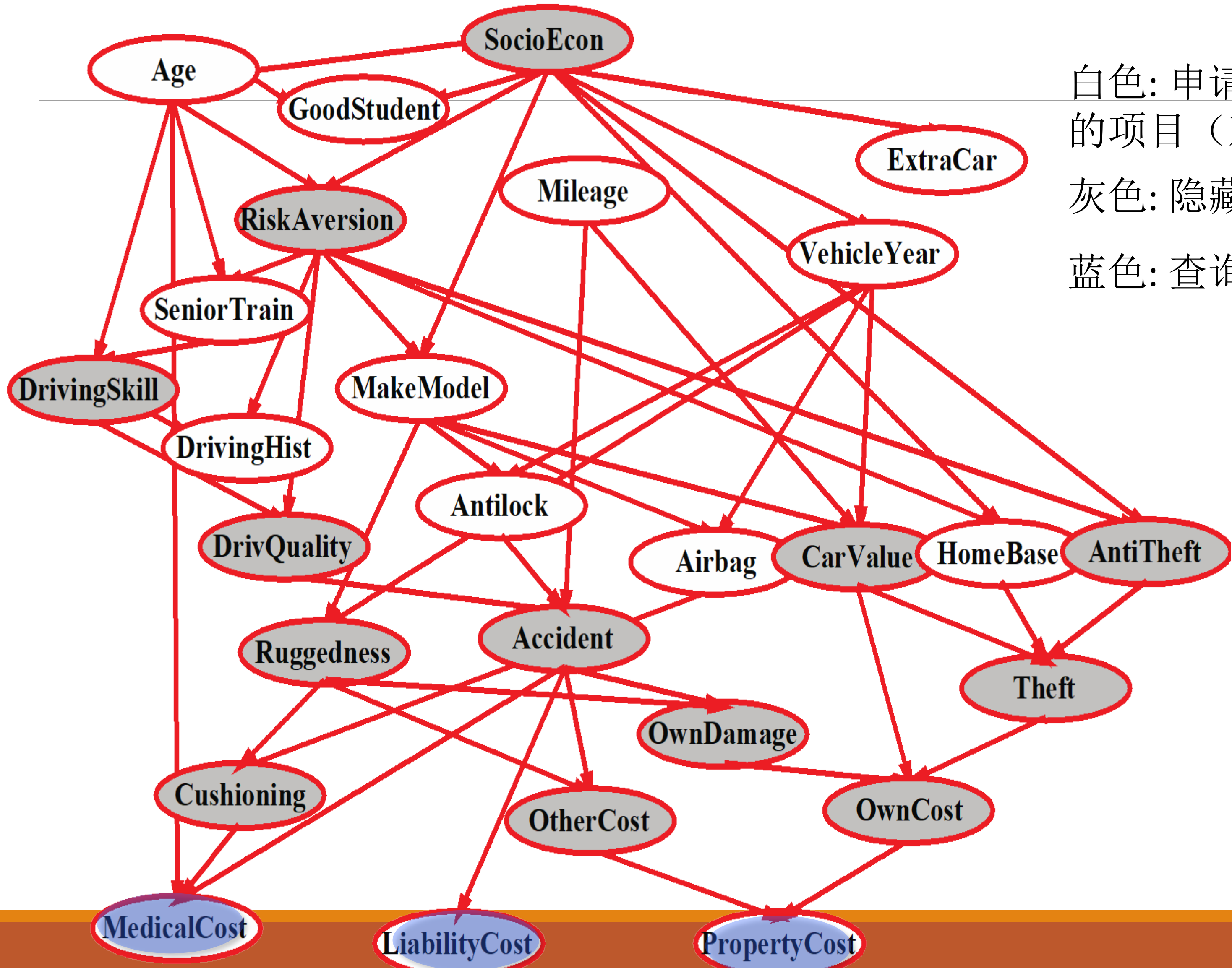
贝叶斯网络: 宏观

- 直接学习 完整联合概率分布表（作为概率模型）的问题:
  - Unless there are only a few variables, the joint is WAY too big to represent explicitly
  - Hard to learn (estimate) anything empirically about more than a few variables at a time （sample complexity）

- 贝叶斯网络: a technique for describing complex joint distributions (models) using simple, local distributions (conditional probabilities)
  - More properly called 概率图形模型
  - We describe how variables locally interact
  - Local interactions chain together to give global, indirect interactions
  - For about 10 min, we'll be vague about how these interactions are specified



贝叶斯网络举例: 汽车保险

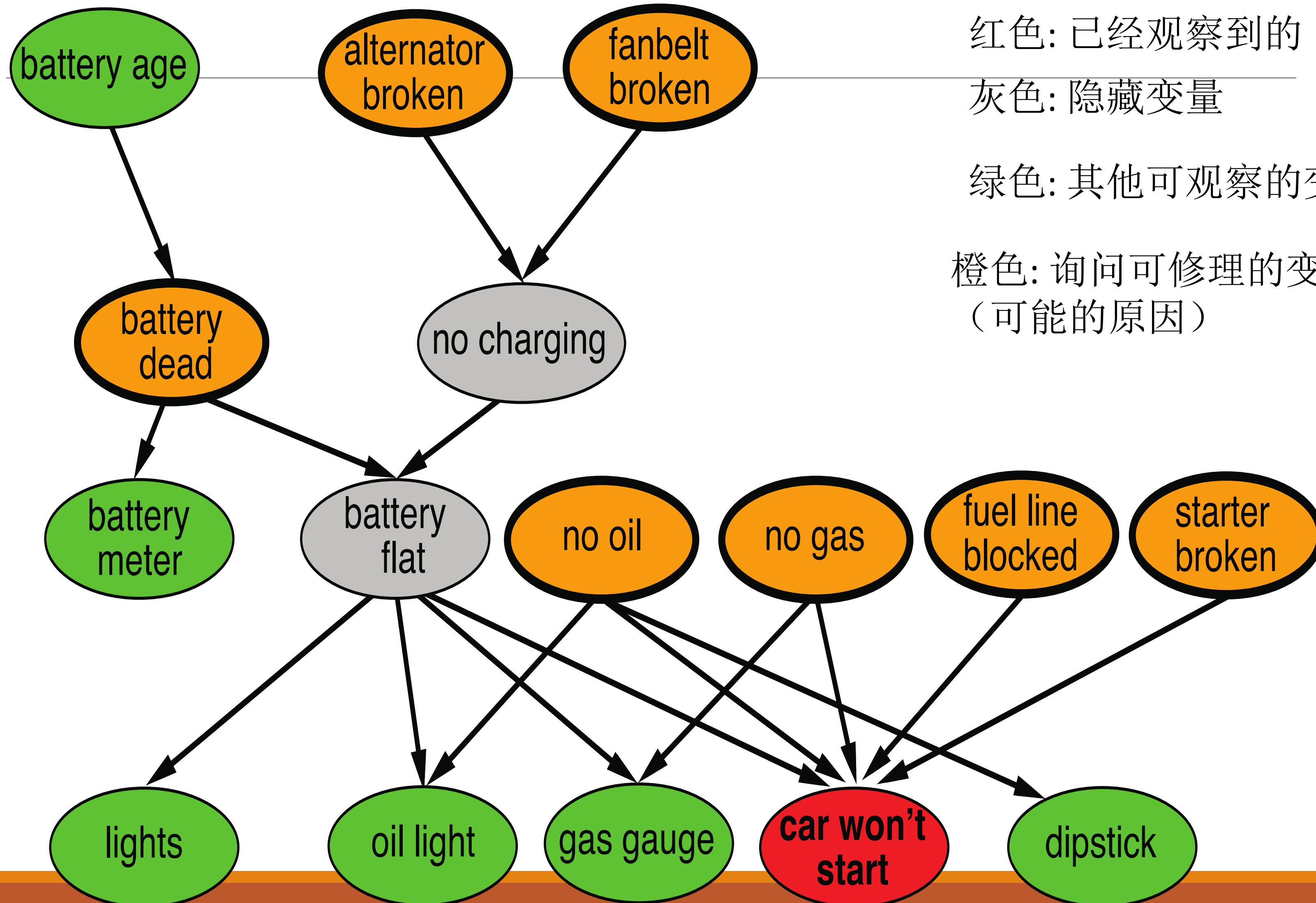


白色: 申请表格里出现的项目（观察到的）

灰色: 隐藏变量

蓝色: 查询变量

贝叶斯网络举例: 简单汽车维修



红色: 已经观察到的

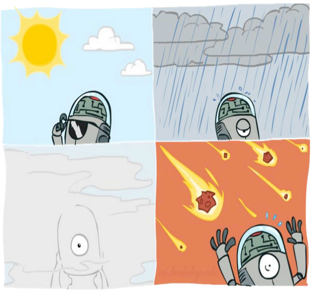
灰色: 隐藏变量

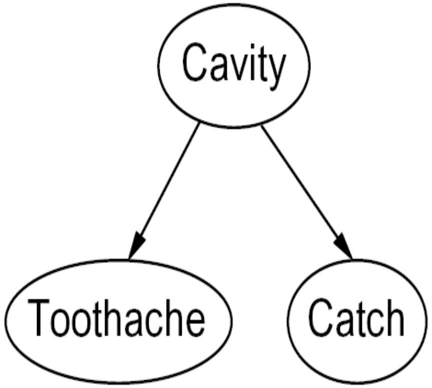
绿色: 其他可观察的变量

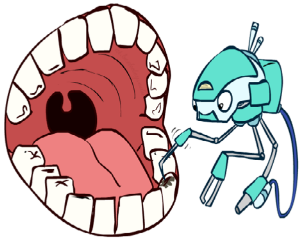
橙色: 询问可修理的变量（可能的原因）

图形模型的表达

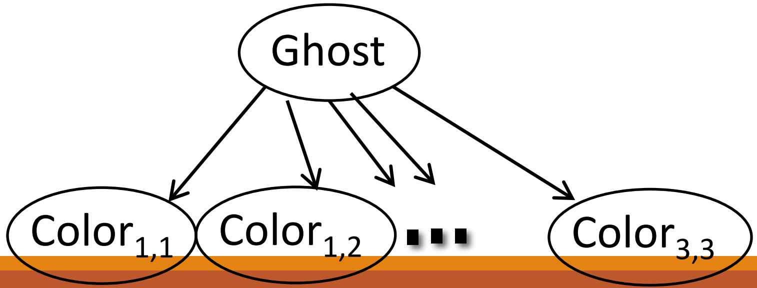
- 节点: 变量 (每个变量有个值域)
- 弧/边: 相互作用
  - 指示 变量间的“直接的影响”
  - 正式的意思: 代表条件独立性
- 现在可以简单地: 箭头意味着直接的因果关系 (通常情况下, 它们并不一定是这样!)





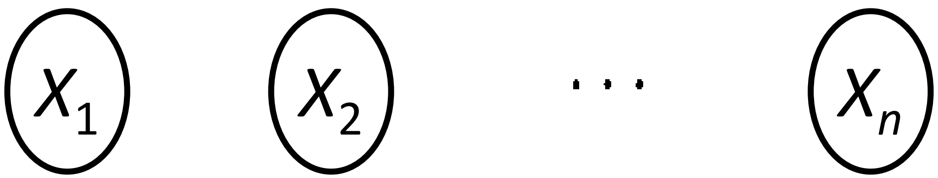


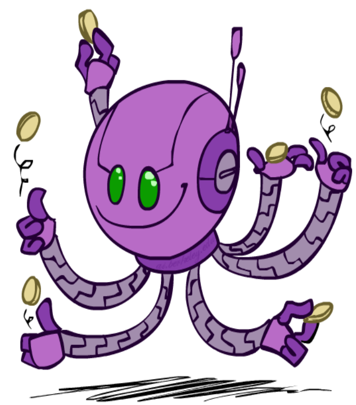
Catch，这里指的是牙医机器人的探测器是否捕获到你的牙洞。



举例: 硬币上抛翻转

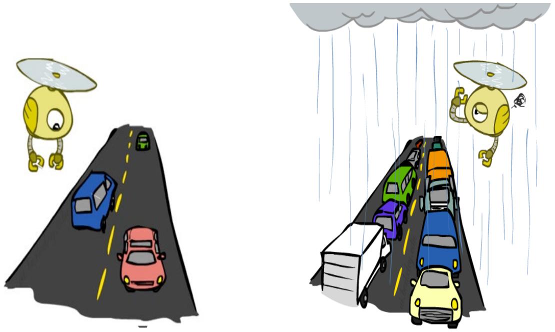
- N 次硬币上抛（也可以理解为N个硬币抛一次）



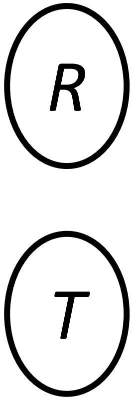


- 变量间没有相互作用: 绝对独立（没有边连接）

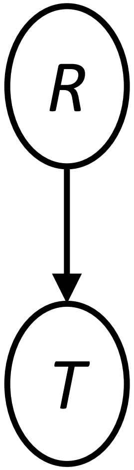
举例: 交通流量预测



- 变量:
  - R: 下雨
  - T: 交通状况
- 模型 1: 独立的



- 模型 2: 下雨导致交通状况变化

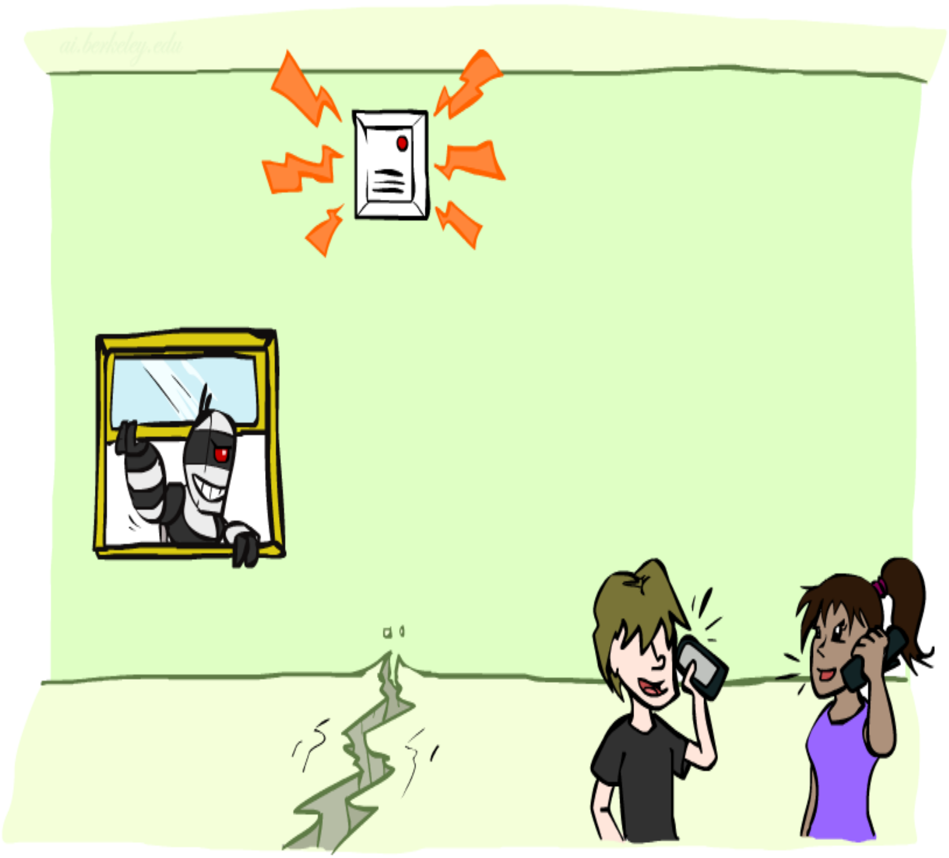


- 为什么用模型 2 更好?

举例: 防盗报警系统

变量

- B: 入室盗窃
- A: 报警器响
- M: 马丽打电话
- J: 约翰打电话
- E: 地震!



贝叶斯网络语法和语义

贝叶斯网络语法

- 一个节点对应一个变量 X_i
- 一个有向, 无环图
- 对每个节点 给定图中它的 *父节点*

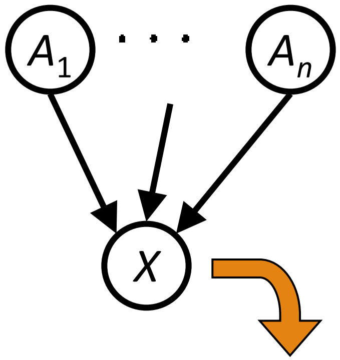
有一个条件概率分布，

  - ***C**PT*: 条件概率分布表:

    在给定父节点的一个配置以后，

    每一行是子节点取值的一个分布 $P(X|a_1 \ldots a_n)$

  - 一个近似的“因果”过程的描述



$P(X|A_1 \ldots A_n)$

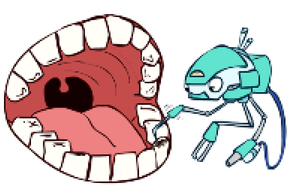
贝叶斯网络 = 拓扑结构(图形) + 局部条件概率

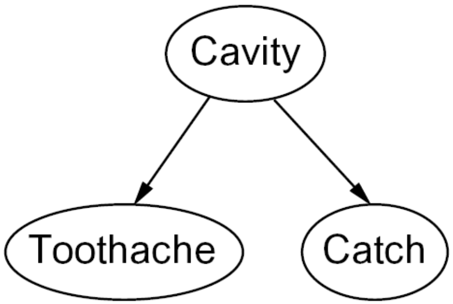
贝叶斯网络里的概率

- 贝叶斯网络 隐式地 蕴含了联合概率分布
  - 局部条件概率分布的乘积
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

  - Example:





$P(+cavity,\ +catch,\ -toothache)$

贝叶斯网络里的概率

- 以下表达式成立，为什么？

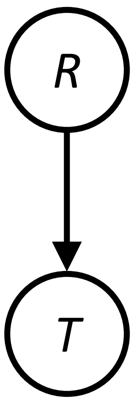
$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$$

- 乘法法则：$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | x_1 \ldots x_{i-1})$
- 假设 条件独立性存在: $P(x_i | x_1, \ldots x_{i-1}) = P(x_i | parents(X_i))$

  → 结果: $P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i | parents(X_i))$

- 不是每一个BN 可以表达与变量相关的任何的联合概率分布
  - 图结构表达了某种特定的条件独立性

举例: Traffic



$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

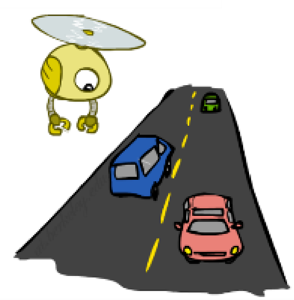
+r

+t	3/4
-t	1/4

-r

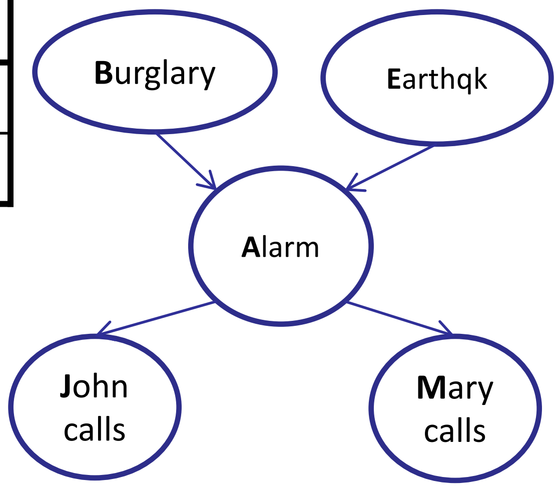
+t	1/2
-t	1/2

$$P(+r, -t) =$$

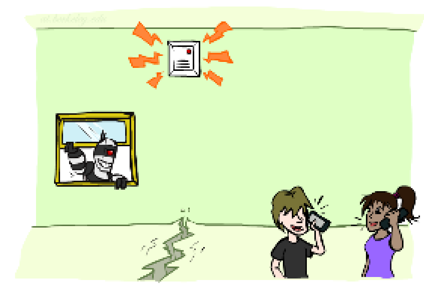


举例: 报警器模型

B	P(B)
+b	0.001
-b	0.999



E	P(E)
+e	0.002
-e	0.998

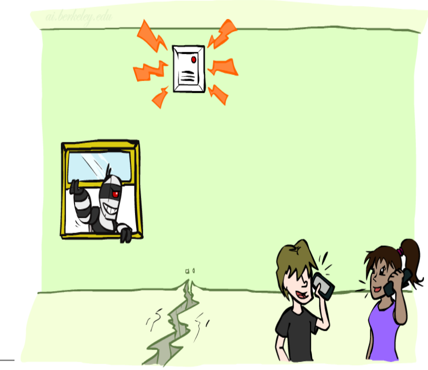


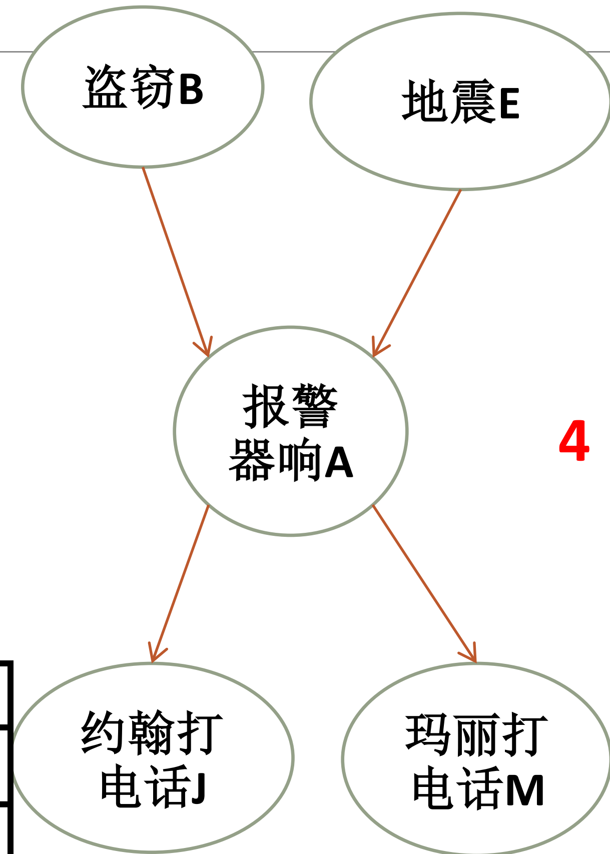
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

举例: 报警器网络





P(B)	
true	false
0.001	0.999

1

P(E)	
true	false
0.002	0.998

1

B	E	P(A|B,E)	
		true	false
true	true	0.95	0.05
true	false	0.94	0.06
false	true	0.29	0.71
false	false	0.001	0.999

4

A	P(J|A)	
	true	false
true	0.9	0.1
false	0.05	0.95

2

A	P(M|A)	
	true	false
true	0.7	0.3
false	0.01	0.99

2

条件概率分布表CPT的自由参数的个数总共有:

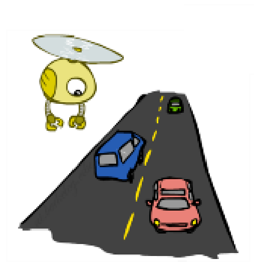
父变量的值域大小: $d_1,\dots,d_k$

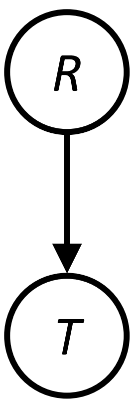
子变量的值域为 d
表中每一行概率值之和为 1

$$(d-1)\prod_i d_i$$

举例: Traffic

- 因果关系方向





$P(R)$

+r	1/4
-r	3/4

$P(T|R)$

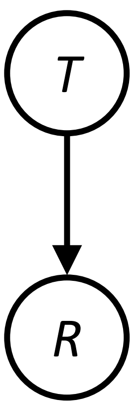
+r	+t	3/4
	-t	1/4
-r	+t	1/2
	-t	1/2

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

举例: Reverse Traffic

- 因果关系反向会怎么样?



$P(T)$

+t	9/16
-t	7/16

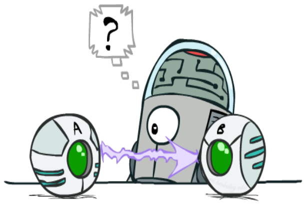
$P(R|T)$

+t	+r	1/3
	-r	2/3
-t	+r	1/7
	-r	6/7

$P(T,R)$

+r	+t	3/16
+r	-t	1/16
-r	+t	6/16
-r	-t	6/16

因果关系(Causality)?



- 当贝叶斯网络反映了真实的因果关系模式时，通常以下成立:
  - 更简单的网络 (较少的父节点, 较少的参数)
  - 更容易评估概率
  - 鲁棒性更强， 比如修改盗窃的频率后应该不影响模型里的其他部分!
- BNs 不需要实际上表达因果关系
  - 有时没有因果网络存在于一个领域 (尤其是在一些变量丢失的情况下)
  - 其结果是箭头关联反映的是*相关性*(correlation), 而不是因果关系
- 箭头实际表示的是什么?
  - 拓扑结构可能碰巧表达的是因果关系
  - 拓扑结构真正表达（编码）的是条件独立性:
    - $P(X_i \mid X_1,\ldots,X_{i-1}) = P(X_i \mid Parents(X_i))$