贝叶斯网络

从贝叶斯定理到贝叶斯网络

高　阳

http://cs.nju.edu.cn/gaoy, 2016.10.12

贝叶斯网络

贝叶斯定理

贝叶斯信念网络

贝叶斯网络的结构学习

贝叶斯网络的参数学习

贝叶斯网络的推理

贝叶斯定理

先验概率 vs 后验概率

◻ 先验概率

- ✓ 在得到任何新证据之前，统计的事件概率。即非条件概率，P(事件)。
- ✓ 一个人脑膜炎的概率，即脑膜炎的人数除以目标区域总人数。

◻ 后验概率

- ✓ 给定新证据之后，统计的事件概率。即条件概率，P(事件|证据)。
- ✓ 一个人表现出头痛的症状，其脑膜炎的概率，P(脑膜炎|头痛)。

演绎推理 vs 归纳推理

演绎推理：不要求前件是真实的

归纳推理：要求前件必须为真，结论未必为真（从特殊到一般）

贝叶斯决策(贝叶斯归纳推理)：

- ✓ 已知先验概率和类条件概率表达式；
- ✓ 转换成后验概率；
- ✓ 根据后验概率大小进行决策/推理。

Thomas Bayes，1701-1761
英国数学家

贝叶斯定理

基本思想：通过先验概率和类条件件概率表达式，计算后验概率

$$P(H|E)=\Phi\{P(E|H)\}$$

推导：

$$P(H|E)=P(H\cap E)/P(E)$$

$$P(E|H)=P(H\cap E)/P(H)$$

$$P(H|E)=\frac{P(E|H)P(H)}{P(E)}$$

一般形式：$P(H_i|E)=\dfrac{P(E|H_i)P(H_i)}{\sum_{k=1}^{n}P(E|H_k)P(H_k)}$

例子

购买汽车时，去第一家经销商d_1的概率为0.2，去第二家经销商d_2的概率为0.4，去第三家经销商d_3的概率为0.4。在经销商d_1处购买汽车a_1的概率为0.2，在经销商d_2处购买汽车a_1的概率为0.4，在经销商d_3处购买汽车a_1的概率为0.3。如果已经购买了汽车a_1，则在d_2处购买的概率是多少？

$$P(d_2 \mid a_1) = \frac{P(a_1 \mid d_2)P(d_2)}{\sum_{k=1}^{3} P(a_1 \mid d_k)P(d_k)} = \frac{0.4 * 0.4}{0.2 * 0.2 + 0.4 * 0.4 + 0.4 * 0.3} = 0.5$$

多个证据下的推理

链式规则（采用归纳法）

✓ 考虑n个证据

$$P(H|E_1 \cap E_2 \cap \cdots \cap E_m) = \frac{P(E_1 \cap E_2 \cap \cdots \cap E_m | H)P(H)}{P(E_1 \cap E_2 \cap \cdots \cap E_m)}$$

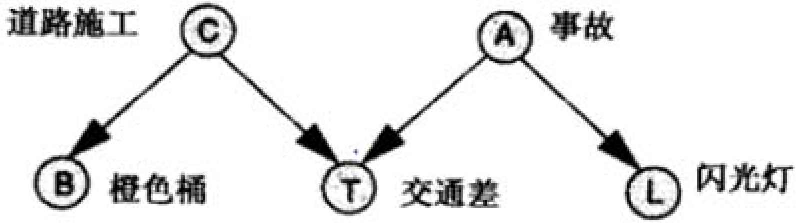
✓ 考虑n-1个证据，直到1个

$$\begin{aligned}
&P(H|E_1 \cap E_2 \cap \cdots \cap E_m) \\
&= \frac{P(E_1 \cap E_2 \cap \cdots \cap E_{m-1}|E_m, H)P(E_m | H)P(H)}{P(E_1 \cap E_2 \cap \cdots \cap E_{m-1}|E_m)P(E_m)} \\
&= \cdots\cdots \\
&= \frac{P(E_1 | E_2 \cdots, E_m, H)P(E_2 | E_3 \cdots, E_m, H) \cdots P(E_m | H)P(H)}{P(E_1|E_2, \cdots, E_m)P(E_2 | E_3, \cdots, E_m) \cdots P(E_{m-1} | E_m)P(E_m)}
\end{aligned}$$

贝叶斯信念网络

另一个例子

假设从仙林开车去鼓楼，发现交通很拥堵。你试图寻找交通差的可能的解释。



C	T	p
t	t	0.3
t	f	0.2
f	t	0.1
f	f	0.4

C is true =0.5

T is true =0.4

计算代价

计算例子中所有参数的联合概率

$$P(C,A,B,T,L) = P(C)P(A\,|\,C)P(B\,|\,A,C)P(T\,|\,B,A,C)P(L\,|\,C,A,B,T)$$

需要存储的联合概率表的表项为$2^5=32$个！

假定这个问题中的参数只与双亲节点的概率相关

$$P(C,A,B,T,L) = P(C)P(A\,|\,C)P(B\,|\,C)P(T\,|\,A,C)P(L\,|\,A)$$

需要存储的联合概率表的表项为$2^2+2^2+2^3+2^2=20$个！

贝叶斯信念网络

◻ 贝叶斯网络

- ✓ 是一个有向无环图。
- ✓ 节点代表随机变量。
- ✓ 边代表节点间的关系（因果关系），用条件概率表达关系的强度。
- ✓ 没有父节点的用先验概率表达信息。

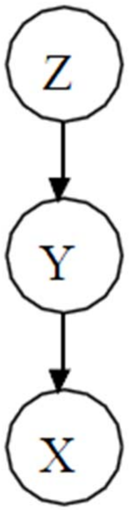
Judea Pearl，1988年提出

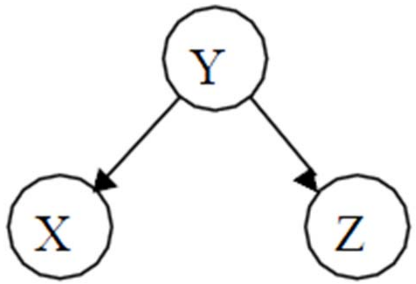
- ✓ 因果关系不可以循环（结果不能推回原因）。
- ✓ 因而，推理就是图中的一条路径。

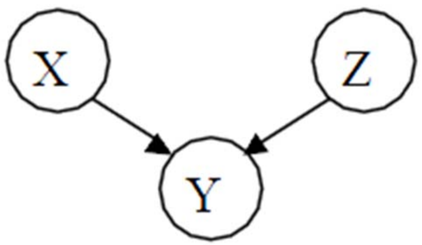
贝叶斯网络中的独立性

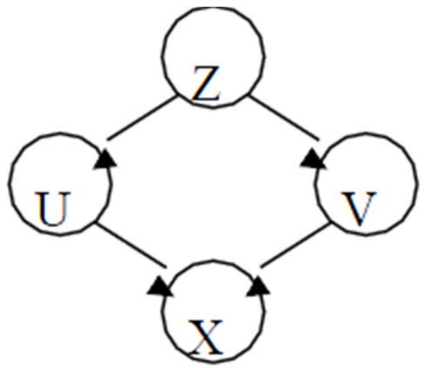
在贝叶斯网络中，条件独立性可以由图的结构判定

- ✓ 顺序连接
- ✓ 分支连接
- ✓ 汇合连接
- ✓ 分支和汇合



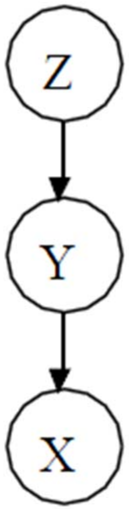






顺序连接

- 顺序连接：给定Y的知识，那么X和Z独立。如果Y未知，则X和Z不独立。



证明：

$$P(Z|Y,X)=\frac{P(X,Y,Z)}{P(X,Y)}$$

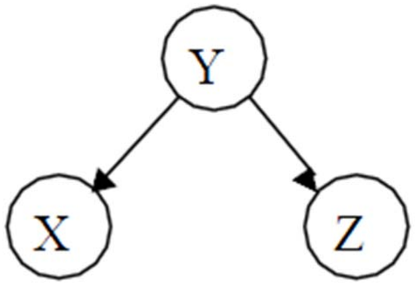
$$=\frac{P(Z)P(Y\,|\,Z)P(X\,|\,Y,Z)}{P(Y)P(X\,|\,Y)}$$

$$=\frac{P(Z)P(Y\,|\,Z)}{P(Y)}=P(Z\,|\,Y)$$

- 油量诊断
  - 旧的活塞(Z)引起油量过度消耗(Y); $P(X,Y,Z)=P(Z)\times P(Y|Z)\times P(X|Y)$
  - 油量过度消耗(Y)导致油箱油量过低(X)。

分支连接

- **分支连接：** 给定Y的知识，那么X和Z独立。Y未知，X和Z不独立。

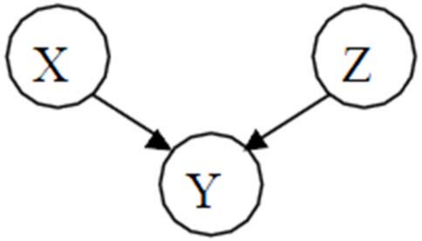


$$P(X,Y,Z)=P(Y)\times P(X|Y)\times P(Z|Y)$$

- 油量诊断
  - 旧的活塞(Y)引起蓝色的废气(X);
  - 旧的活塞(Y)引起油量过度消耗(Z); 。

汇合连接

- **汇合连接：** Y未知时，X和Z独立。但给定Y，X和Z不独立。

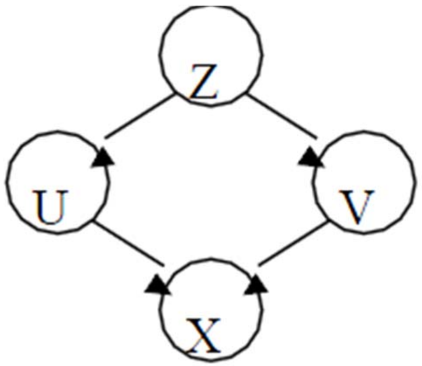


$$P(X,Y,Z)=P(X)\times P(Z)\times P(Y|X,Z)$$

- 油量诊断
  - 油量过度消耗(X)导致油箱油量过低(Y)；
  - 漏油(Z)导致油箱油量过低(Y) 。

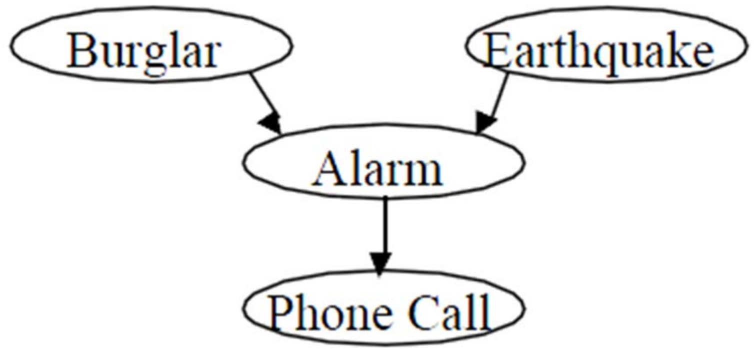
分支和汇合

- **分支和汇合：** 仅给定U，Z和X不独立。当且仅当给定U和V时，Z和X独立。



汇合连接的独立性

- 防盗警报响可能是窃贼引起的；
- 防盗警报响也可能是地震引起的；
- 窃贼和地震是相互独立的；
- 假设，你听到了警报响，并已知附近发生了轻微地震……
- 那么，你会推理，警报响是由于地震引起而不是由窃贼…..



因此，在给定Alarm或其子节点，

Earthquake和Burglar不独立！

d-可分

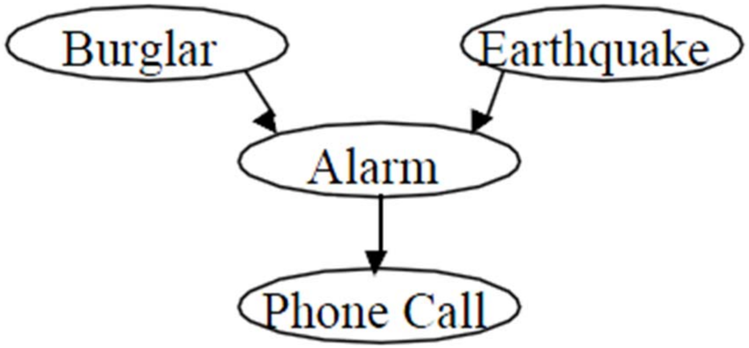
判断贝叶斯网络中任意两个节点之间是否独立

定义： A和B被一组随机变量E d-可分，当且仅当他们之间的所有路径都是堵塞的。

堵塞： 如果A到B上有这样的一个中间节点V，那么路径是堵塞的。V满足以下两个属性之一：

（一）连接是顺序的或者分支的，V在E中。

（二）连接是汇合的，则V和它的子节点都不在E中。



d-可分

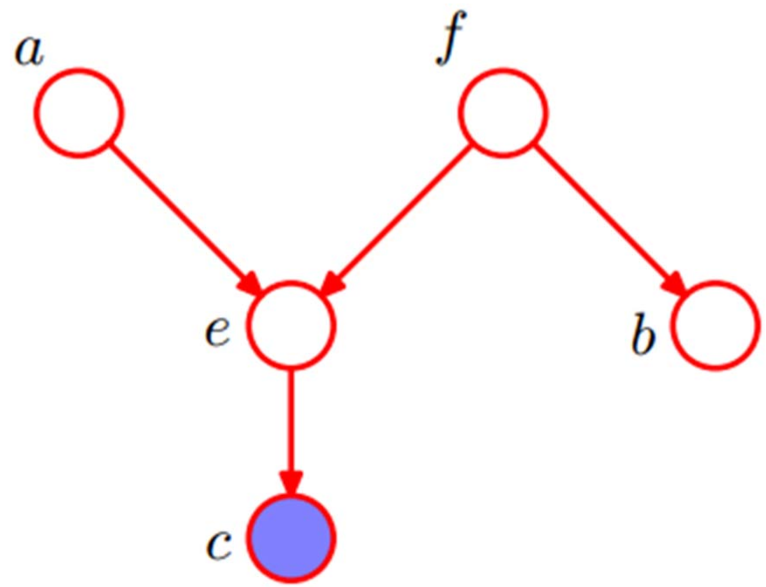
问题：a与b是否在c条件下独立？

路径：a到b的路径为a→e→f→b,考虑e,f的性质

（一）f是分支连接类型。f未知，路径没有被阻断。

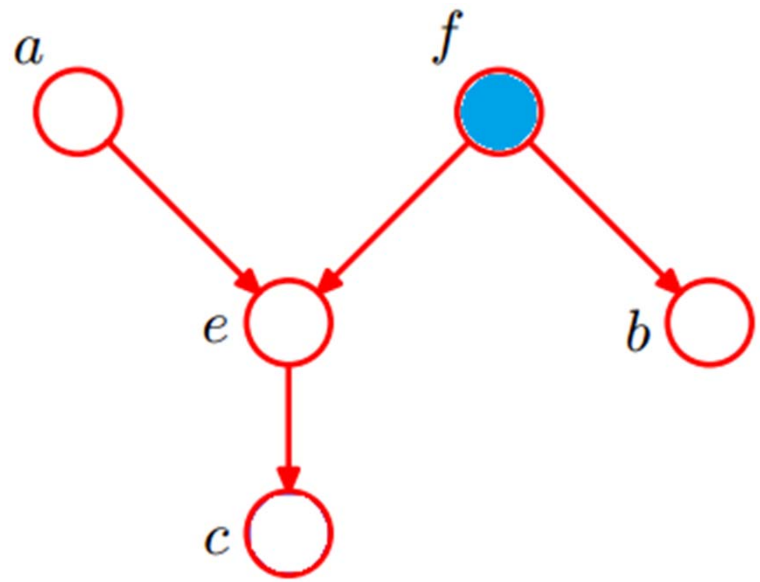
（二）e是汇合连接类型的。e的后继节点c已知，路径没有被阻断。

因此：a与b在c条件下不独立



d-可分

问题：a与b是否在f条件下独立？



路径：a到b的路径为a→e→f→b,考虑e,f的性质

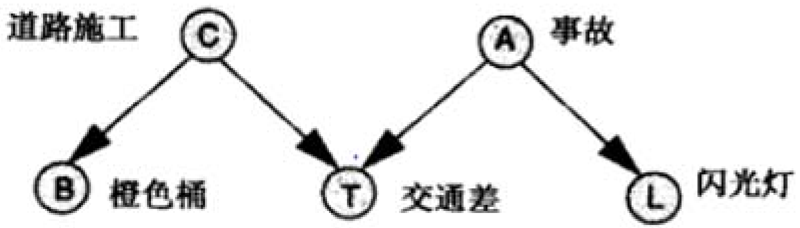
（一）f是分支连接类型。f已知，路径被阻断。

（二）e是汇合连接类型的。e的后继节点c未知，路径被阻断。

因此：a与b在f条件下独立

例子的再考察

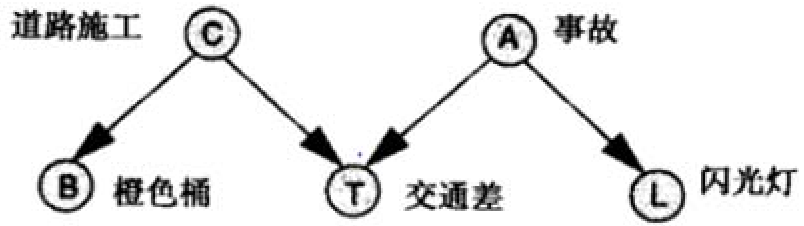
假设从仙林开车去鼓楼，发现交通很拥堵。你试图寻找交通差的可能的解释。



$$P(C,A,B,T,L)=P(C)P(A\mid C)P(B\mid A,C)P(T\mid B,A,C)P(L\mid C,A,B,T)$$

$$P(C,A,B,T,L)=P(C)P(A)P(B\mid C)P(T\mid A,C)P(L\mid A)$$

例子的再考察



$$P(C,A,B,T,L)=P(C)P(A\,|\,C)P(B\,|\,A,C)P(T\,|\,B,A,C)P(L\,|\,C,A,B,T)$$

◘ 用d-可分化简

- ✓ 对于第二项P(A|C),其是汇合连接。在T未知下，A和C相互独立。因此P(A|C)=P(A)。
- ✓ 对于第三项P(B|A,C),在B到A的路径上有T。T是汇合连接。在T未知下，路径堵塞，B和A相互独立。因此P(B|A,C)=P(B|C)。
- ✓ 对于第四项P(T|B,A,C),在T到B的路径上有C。如果C已知，路径阻断。T,B在C已知情况下条件独立。因此P(T|B,A,C)=P(T|A,C)。
- ✓ 对于第五项P(L|T,B,A,C),在L到B,C,T的路径上有A。如果A已知，路径阻断。L分别与B,C,T在A已知情况下条件独立。因此P(L|T,B,A,C)=P(L|A)。

$$P(C,A,B,T,L)=P(C)P(A)P(B\,|\,C)P(T\,|\,A,C)P(L\,|\,A)$$

贝叶斯网络的构建

□定义变量

✓在领域专家指导下选取合适变量，或从中选择重要的因子。

□结构学习

✓构建有向无环图。

✓能够很好的解释数据，反应变量间的依赖关系或者独立性。

✓不造成过拟合。

□参数学习

✓学习节点的分布参数，即每条边对应的条件概率分布。

网络结构确定

- 选择一组刻画问题的随机变量$\{X_1, X_2, \ldots, X_n\}$
- 确定一个变量顺序$a=< X_1, X_2, \ldots, X_n >$
- 参数学习从一个空图出发，按照顺序a逐个将变量加入ξ中
- 假设当前需要加入的是变量X_i，此时ξ中已包含变量$X_1, X_2, \ldots, X_{i-1}$
  - ✓ 在$X_1, X_2, \ldots, X_{i-1}$选择一个尽可能小的子集$\pi(X_i)$，使得假设“给定$\pi(X_i)$，$X_i$与$\xi$中的其他变量条件独立”合理
  - ✓ 从$\pi(X_i)$中的每一个节点添加一条指向X_i的有向边

例子回顾-警报

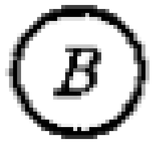
- 防盗警报响A可能是窃贼B引起的；
- 防盗警报响A也可能是地震E引起的；
- 窃贼和地震是相互独立的；
- 听到了警报响，邻居Mary会打电话；
- 听到了警报响，邻居John会打电话；

定义序为a_1=<B,E,A,M,J>

例子回顾–警报

定义序为a_1=<B,E,A,M,J>

- 将B加入空图；



(a) $\mathcal{G}_1$

例子回顾-警报

定义序为a_1=<B,E,A,M,J>

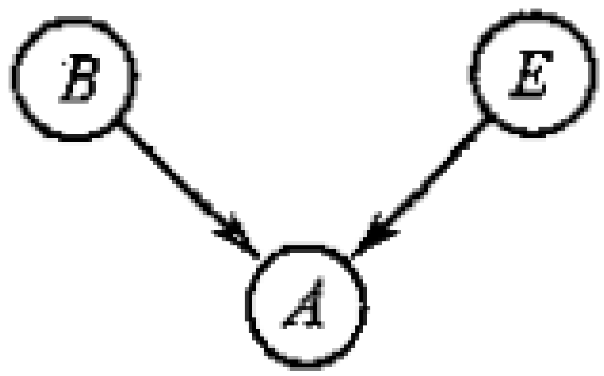
- 将B加入空图；
- 加入E。$\pi(E)=\Phi$

(b) $\mathcal{G}_2$

例子回顾–警报

定义序为a_1=<B,E,A,M,J>

- 将B加入空图；
- 加入E。$\pi(E)=\Phi$
- 加入A。$\pi(A)=\{B,E\}$

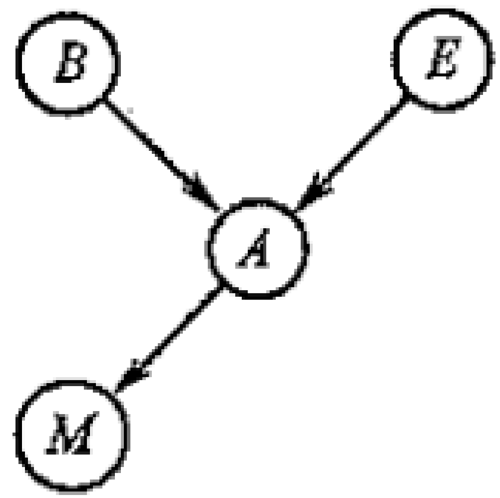


(c) $\mathcal{G}_3$

例子回顾–警报

定义序为a_1=<B,E,A,M,J>

- 将B加入空图；
- 加入E。$\pi(E)=\Phi$
- 加入A。$\pi(A)=\{B,E\}$
- 加入M。$\pi(M)=\{A\}$

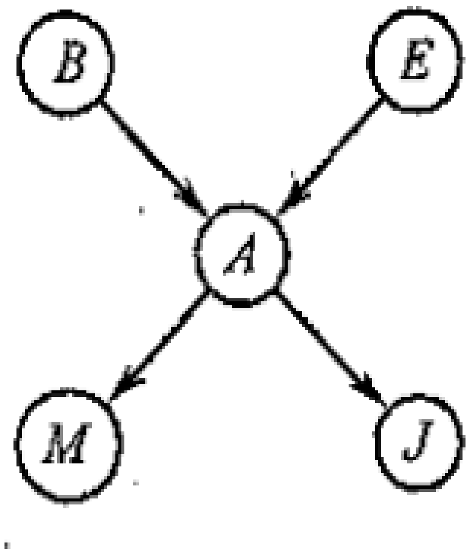


(d) $\mathcal{G}_4$

例子回顾-警报

定义序为a_1=<B,E,A,M,J>

- 将B加入空图；
- 加入E。$\pi(E)=\Phi$
- 加入A。$\pi(A)=\{B,E\}$
- 加入M。$\pi(M)=\{A\}$
- 加入J。$\pi(J)=\{A\}$

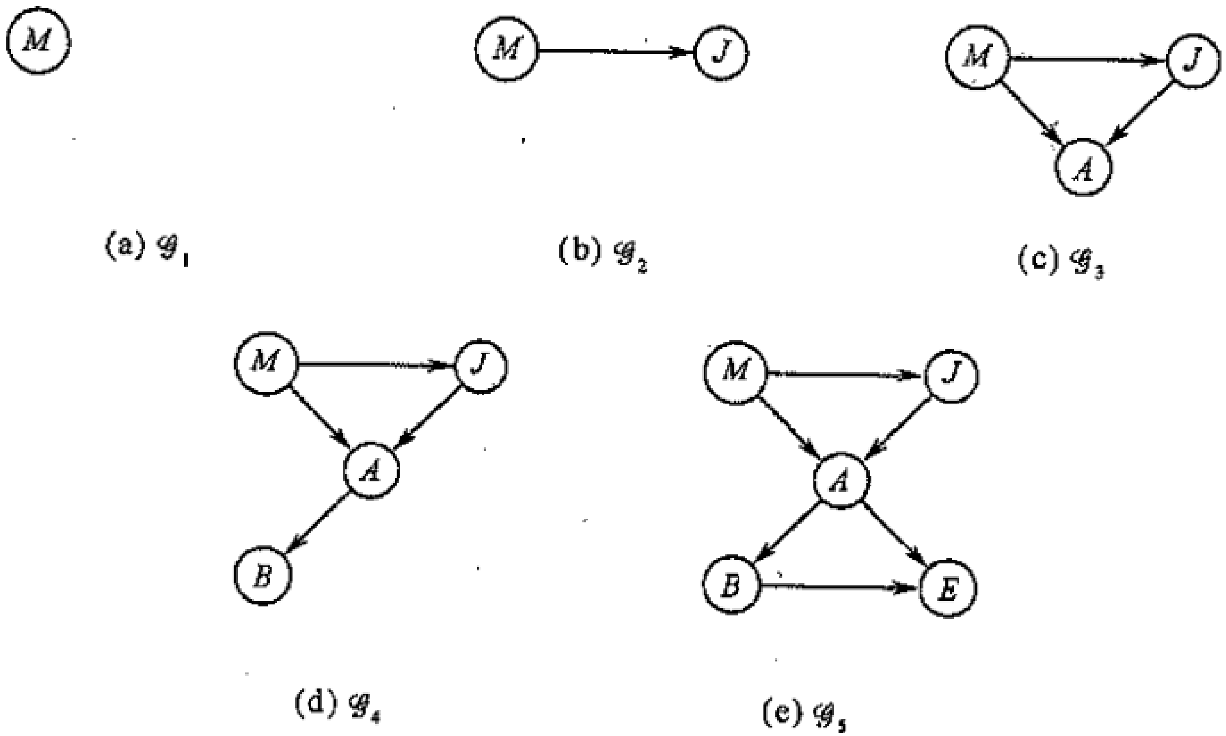


(e) $\mathcal{G}_5$

序的影响

不同的序产生不同的网络结构

定义序为$a_2=<M,J,A,B,E>$，并假定变量出现的次序代表因果关系



(a) $\mathcal{G}_1$ (b) $\mathcal{G}_2$ (c) $\mathcal{G}_3$

(d) $\mathcal{G}_4$ (e) $\mathcal{G}_5$

贝叶斯网络的结构学习

结构学习

结构学习

- 在数据中推断变量之间的依赖关系，在可能的结构空间中搜索最优结构（基于专家的结构学习 vs 基于数据的结构学习）
- 基于搜索和评分的方法
  - ✓ 评分函数
  - ✓ 空间搜索策略
  - ✓ 基于评分函数，搜索与样本数据匹配程度最高的网络结构
- 基于约束的方法

结构学习

- 利用训练样本集，尽可能结合先验知识，确定和训练样本集合匹配最好的贝叶斯网络结构
- 对于有n个变量，可能的结构数目

$$f(n)=\sum_{i=1}^{n}(-1)^{i+1}\frac{n!}{i!(n-i)!}2^{i(n-1)}f(n-i)$$

- 结构学习是一个典型的NP难问题

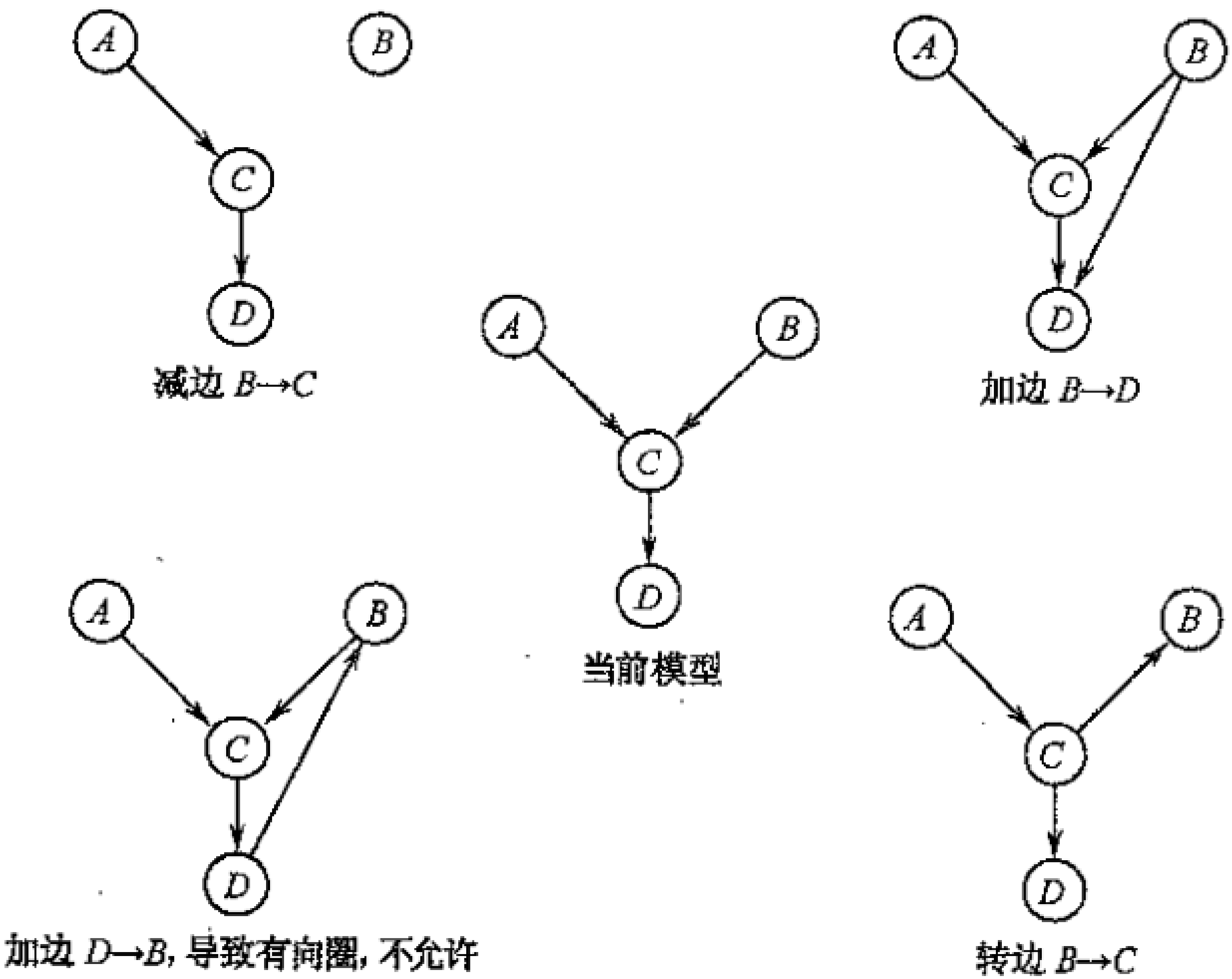
基于搜索和评分的方法

- 利用评分函数，寻找与训练样本匹配最好的贝叶斯网络结构

$$G^* = \arg\max_{G \in \xi} g(G:D)$$

- 贪婪算法
  - 从一个特定的网络出发（如一个没有任何连接边的网络）
  - 利用搜索算法对网络进行操作（增加边，删除边，反转边的方向）
  - 根据评分函数对网络进行评分
  - 检查新的网络结构是否优于旧的，如是，则继续……

搜索算子



评分函数和搜索策略

❑ 评分函数

- ✓ 最早是由Cooper and Herskovits等人在1992年提出的K2评分函数，该评分函数假设观测到的数据是完备的，且服从多项式分布
- ✓ 基于K2评分函数，Heckerman等人在1995年提出了BD评分函数，该评分函数假设观测数据服从Dirichlet分布

❑ 搜索策略

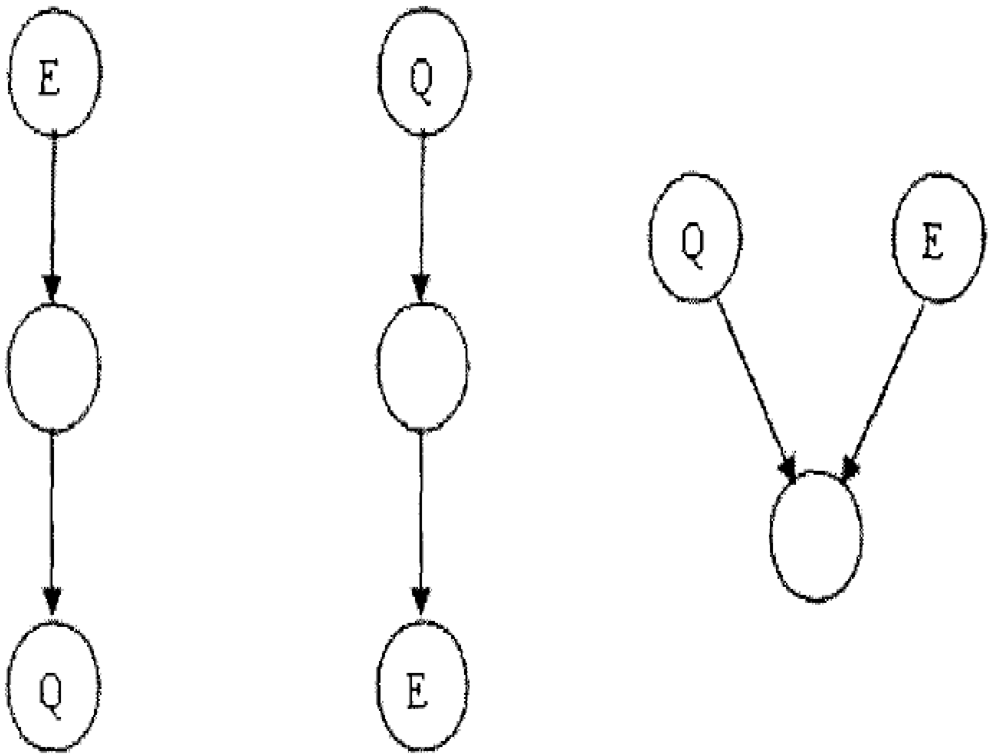
- ✓ 贪婪搜索、模拟退火、禁忌搜索、遗传算法等

贝叶斯网络的推理

贝叶斯信念网络推理

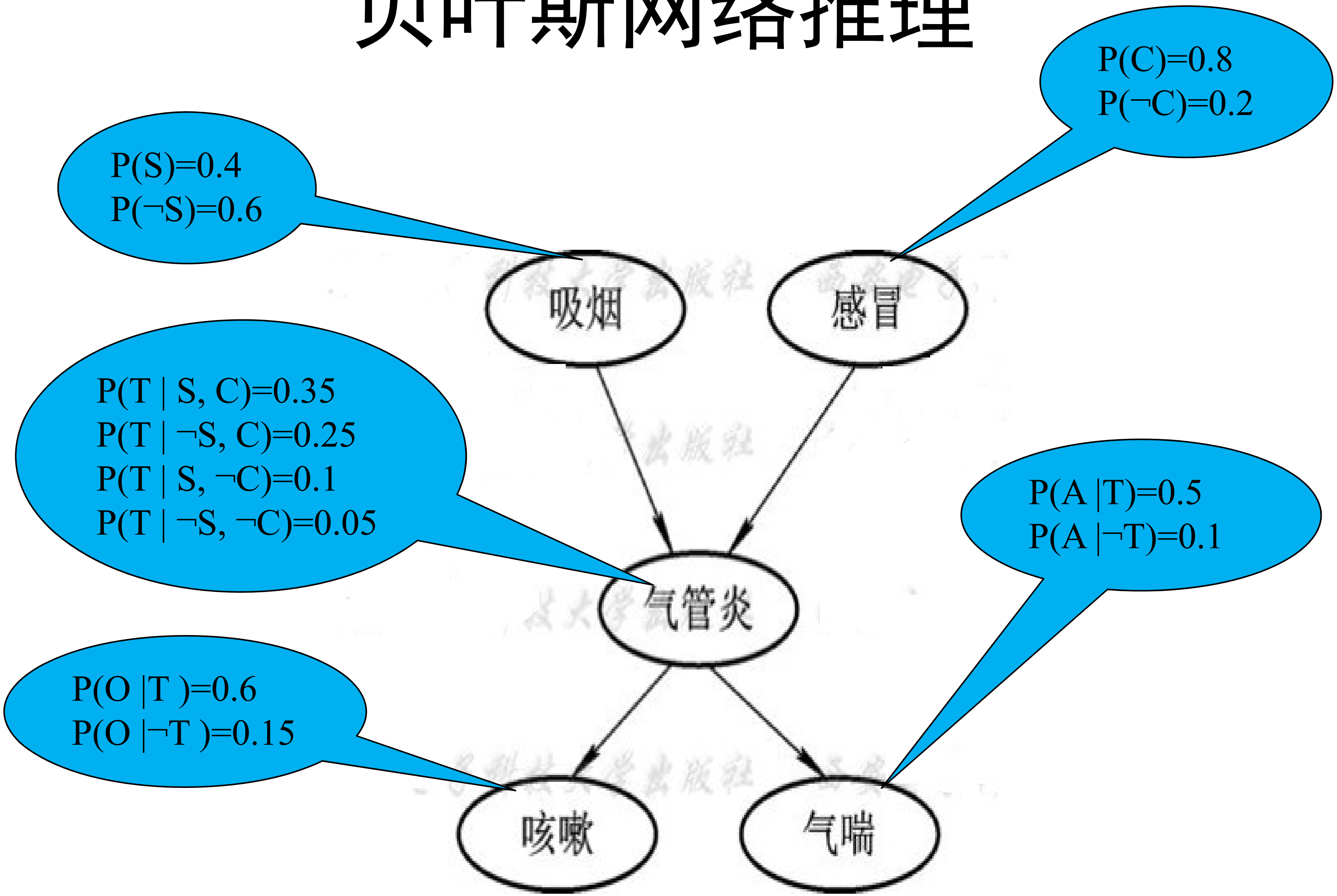
- 因果推理（自顶向下的推理）
  - ✓由原因推出结论，即根据一定的原因，推理出在该原因情况下结果发生的概率。
- 诊断推理（自底向上的推理）
  - ✓由结论推出原因，即根据产生的结果，利用贝叶斯网推理算法，得出导致该结果的原因的概率。
- 支持推理
  - ✓对所发生的现象提供解释，目的是分析原因之间的相互影响。

贝叶斯信念网络推理



因果推理　诊断推理　支持推理

贝叶斯网络推理



P(S)=0.4
P(¬S)=0.6
P(C)=0.8
P(¬C)=0.2
吸烟
感冒
P(T | S, C)=0.35
P(T | ¬S, C)=0.25
P(T | S, ¬C)=0.1
P(T | ¬S, ¬C)=0.05
气管炎
P(A |T)=0.5
P(A |¬T)=0.1
P(O |T)=0.6
P(O |¬T)=0.15
咳嗽
气喘

因果推理

- ✓ 已知网络中的祖先节点而计算后代节点的条件概率
- ✓ 假设已知某人吸烟(S), 计算他患气管炎(T)的概率 $P(T|S)$
- ✓ 由于T还有另一个因节点感冒(C), 对概率 $P(T|S)$ 进行扩展

$$P(T\,|S) = P(T, C\,|S) + P(T, \neg C\,|S)$$

因果推理

✓扩展第一项

$P(T, C \mid S) = P(T, C, S)/P(S)$ (对$P(T, C \mid S)$逆向使用概率的乘法公式)

$= P(T \mid C, S)P(C, S)/P(S)$ (对$P(T, C, S)$使用乘法公式)

$= P(T \mid C, S)P(C \mid S)$ (对$P(C, S)/P(S)$使用乘法公式)

$= P(T \mid C, S)P(C)$ (因为C与S条件独立)

✓同理

$$P(T, \neg C \mid S) = P(T \mid \neg C, S)P(\neg C)$$

因果推理

- ✓代入

$$P(T|S) = P(T|C, S)P(C) + P(T|\neg C, S)P(\neg C)$$

- ✓等式右端的概率值在CPT中已给出，即都为已知

$$P(T|S) = 0.35 * 0.8 + 0.1 * 0.2 = 0.3$$

- ✓即吸烟可引起气管炎的概率为0.3。

因果推理解题方法

- ✓ 首先，对于所求的询问节点的条件概率，用所给证据节点和询问节点的所有因节点的联合概率进行重新表达。
- ✓ 然后，对所得表达式进行适当变形，直到其中的所有概率值都可以从问题贝叶斯网络的CPT中得到。
- ✓ 最后，将相关概率值代入概率表达式进行计算即得所求询问节点的条件概率。

诊断推理

- 已知网络中的后代节点而计算祖先节点的条件概率
- 假设已知某人患了气管炎(T), 计算他吸烟(S)的后验概率P(S |T)。

$$P(S|T)=\frac{P(T|S)P(S)}{P(T)}$$

诊断推理

由前面的因果推理可知

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)}$$

$$P(T \mid S) = 0.3$$

由条件概率表 P(S)=0.4

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} = \frac{0.3 \times 0.4}{P(T)} = \frac{0.12}{P(T)}$$

由因果推理方法有

$$P(T|\neg S) = P(T, C \mid \neg S) + P(T, \neg C \mid \neg S)$$

$$= P(T \mid C, \neg S)P(C) + P(T \mid \neg C, \neg S)P(\neg C)$$

$$= 0.25 * 0.8 + 0.2 * 0.2 = 0.21$$

诊断推理

由因果推理方法有

$$P(T|\neg S) = P(T, C |\neg S) + P(T, \neg C |\neg S)$$

$$= P(T |C, \neg S)P(C) + P(T |\neg C, \neg S)P(\neg C)$$

$$= 0.25 * 0.8 + 0.05 * 0.2 = 0.21$$

$$P(\neg S|T) = \frac{P(T|\neg S)P(\neg S)}{P(T)} = \frac{0.21 \times 0.6}{P(T)} = \frac{0.126}{P(T)}$$

$P(S|T)+P(\neg S|T)=1$ $\qquad \frac{0.12}{P(T)} + \frac{0.126}{P(T)} = 1$ $\qquad P(T) = 0.246$

诊断推理

- ✓ 则

$$P(T|S)=\frac{P(T|S)P(S)}{P(T)}=\frac{0.12}{0.246}=0.4878$$

- ✓ 即该人的气管炎是由吸烟导致的概率为0. 4878

诊断推理的一般思路和方法

- ✓ 先利用贝叶斯公式将诊断推理问题转化为因果推理问题；再用因果推理的结果，导出诊断推理的结果

思考和讨论

- ❑ 贝叶斯网络的近似推理
- ❑ 连接树

专题报告

1. 贝叶斯网络结构学习中的K2评分函数。

要　求：报告时长30分钟，留10分钟讨论。

报告人：周浩

2. 贝叶斯网络结构学习中的BD评分函数。

要　求：报告时长30分钟，留10分钟讨论。

报告人：李潇

专题报告

3. 基于约束的贝叶斯网络结构学习方法。

要　求：报告时长30分钟，留10分钟讨论。

报告人：施呈韵

4. 贝叶斯网络的参数学习。

要　求：报告时长30分钟，留10分钟讨论。

报告人：姜允执

谢 谢 ！