1. 逻辑与推理(2)

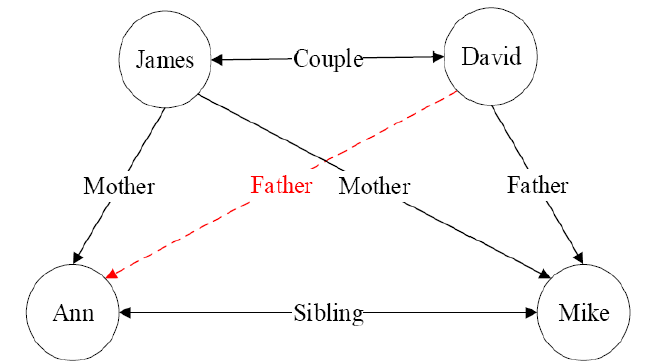
命题逻辑

谓词逻辑

知识图谱推理：一阶归纳推理算法

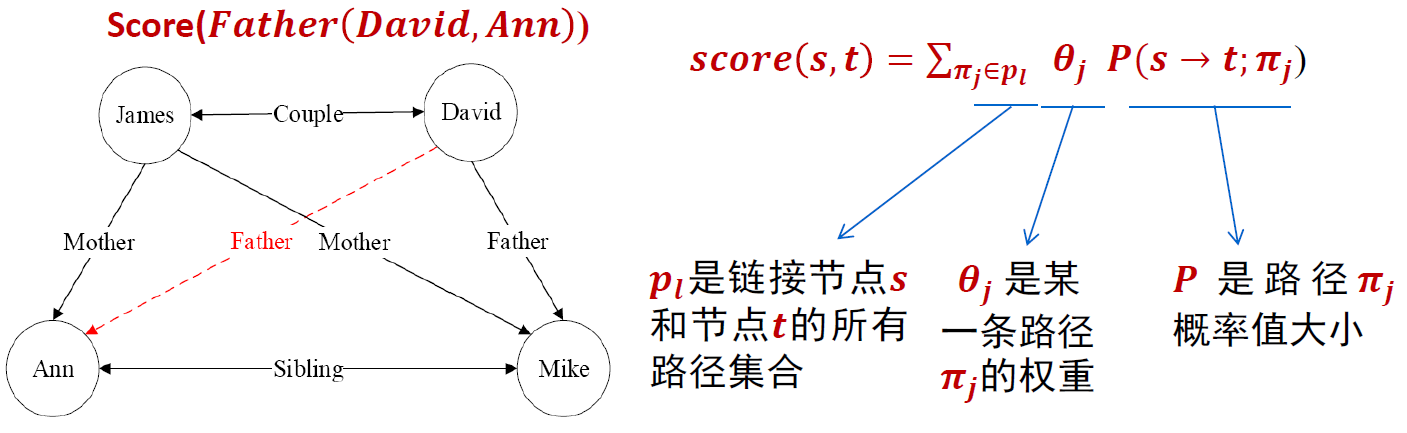
因果推理

2. 路径排序算法



将实体之间的关联路径作为特征，来学习目标关系的分类器

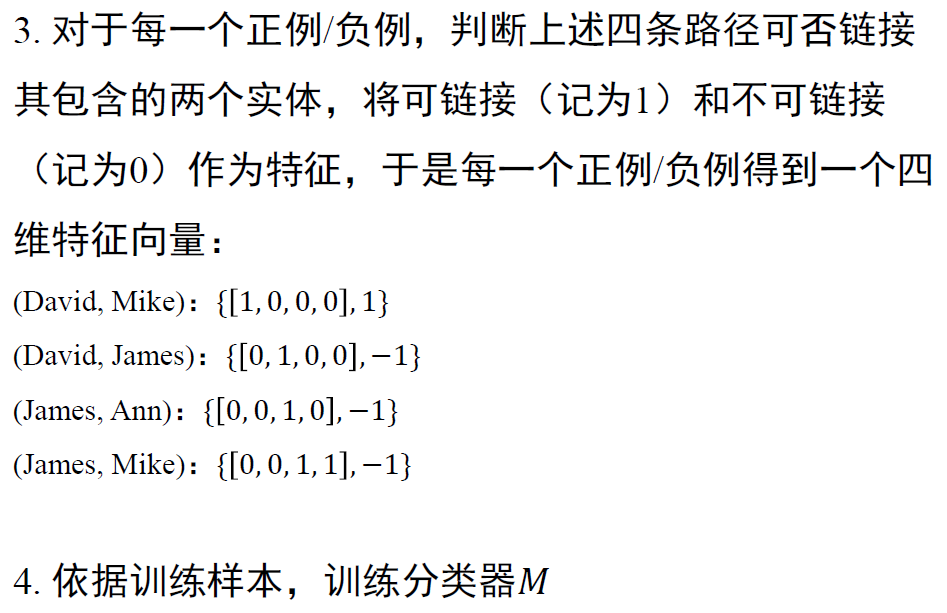
即：判断David和Ann之间的路径关联是否足够支持表述𝑭𝒂𝒕𝒉𝒆𝒓这一关系。

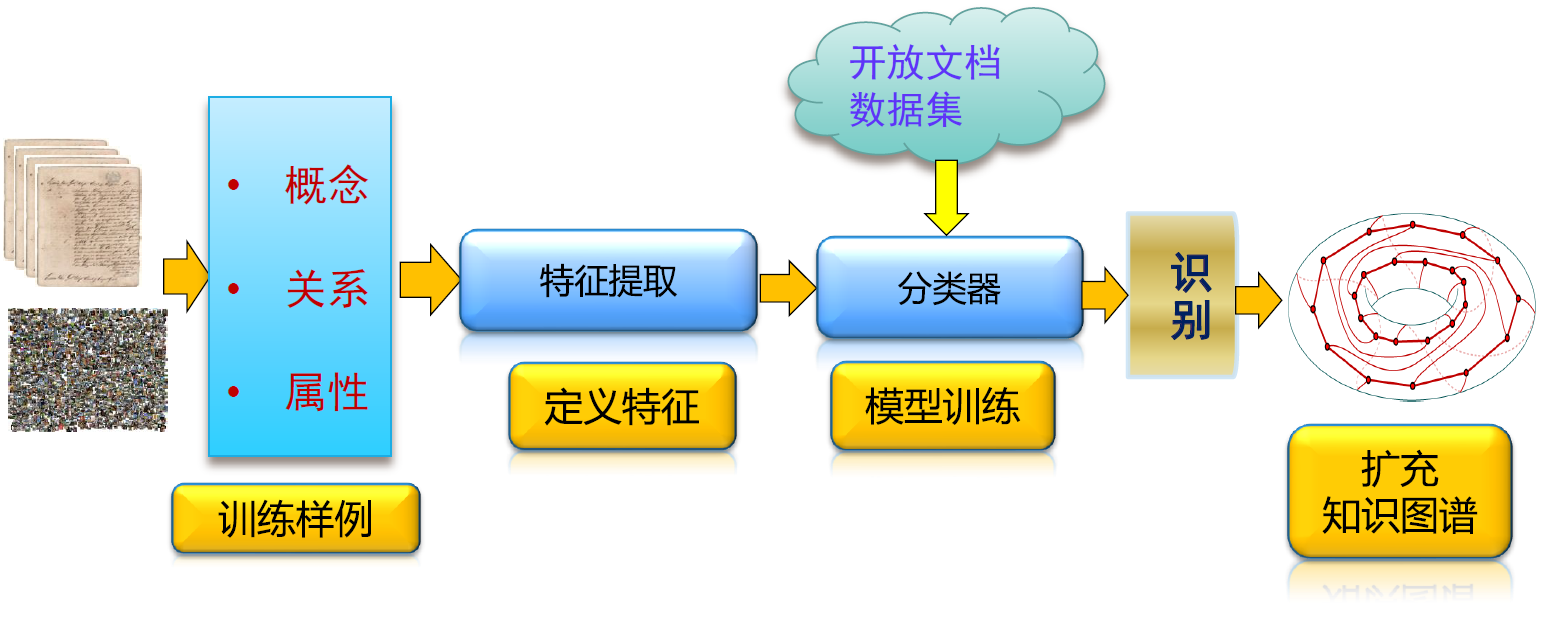


🟎特征抽取：生成并选择路径特征集合。生成路径的方式有随机游走（random walk）、广度优先搜索、深度优先搜索等。

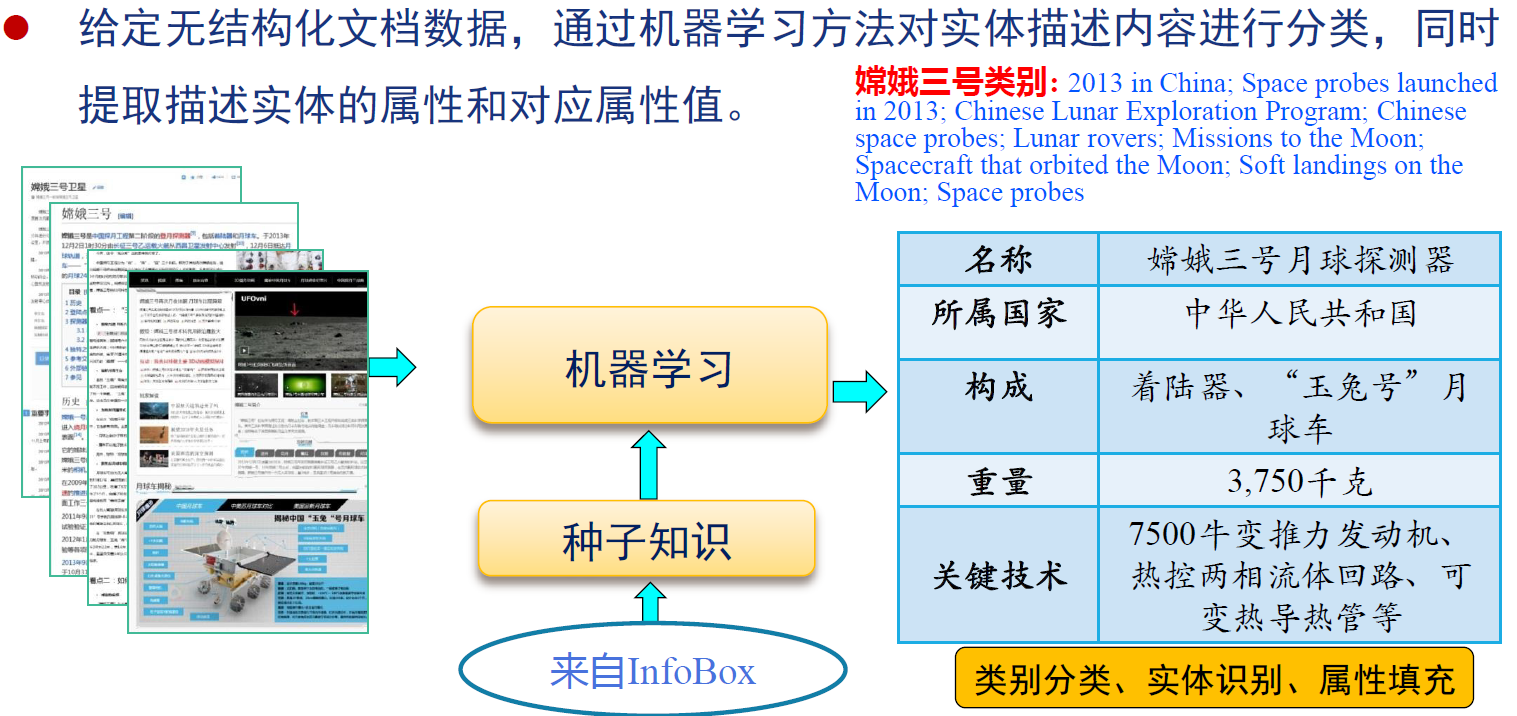
🟎特征计算：计算每个训练样例的特征值𝑃(𝑠→𝑡;𝜋𝑗)。该特征值可以表示从实体节点𝑠出发，通过关系路径𝜋𝑗到达实体节点𝑡的概率；也可以表示为布尔值，表示实体𝑠到实体𝑡之间是否存在路径𝜋𝑗；还可以是实体𝑠和实体𝑡之间路径出现频次、频率等。

🟎分类器训练：根据训练样例的特征值，为目标关系训练分类器。当训练好分类器后，即可将该分类器用于推理两个实体之间是否存在目标关系。

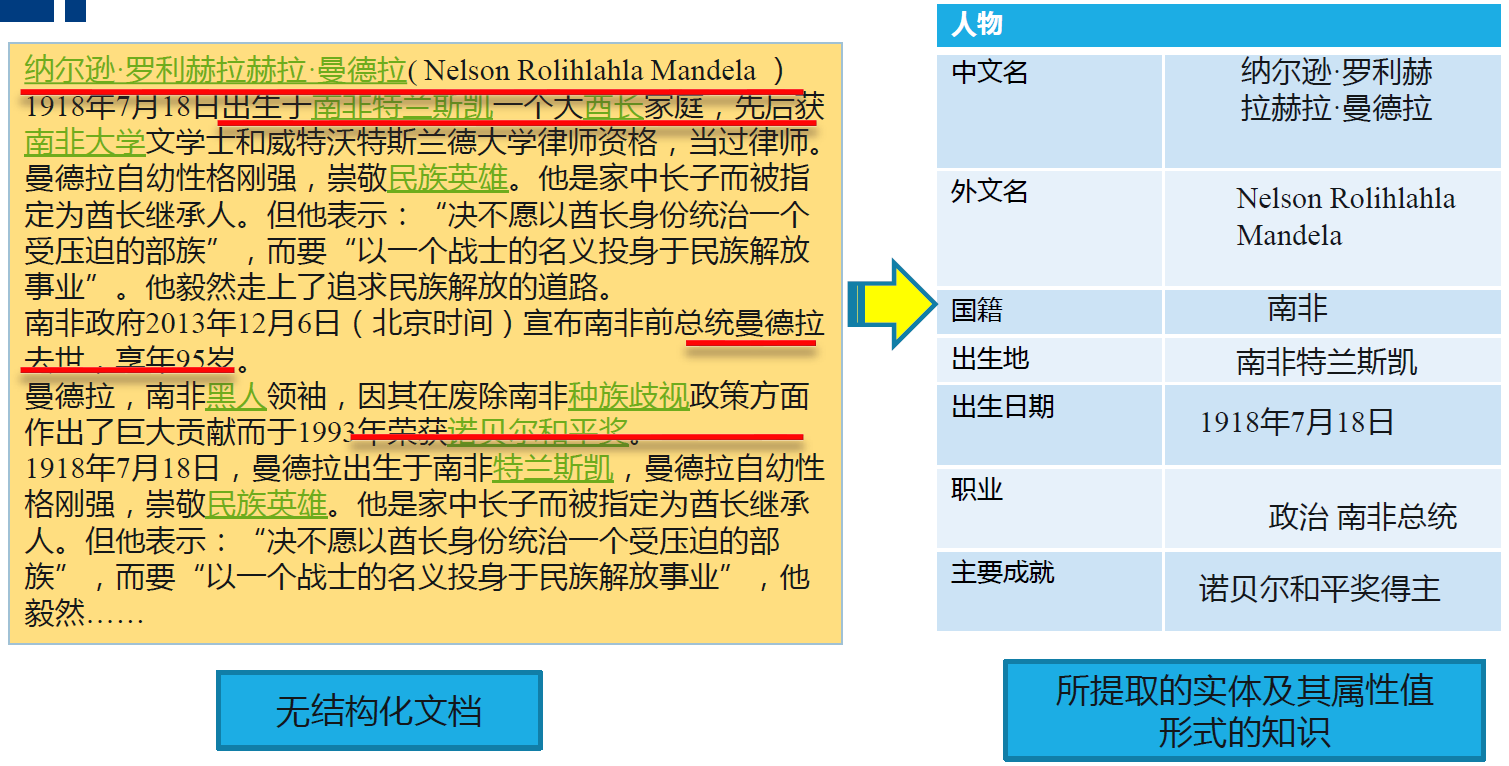
 

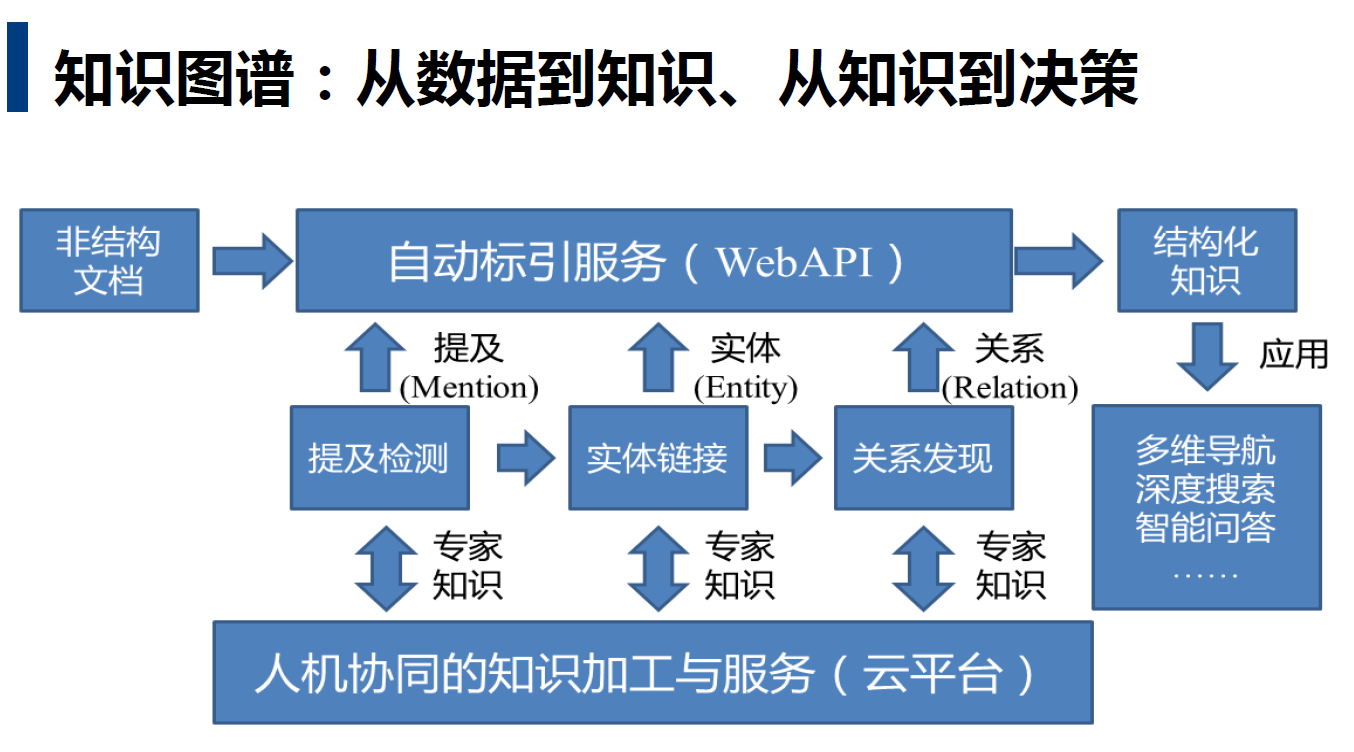


3. 知识图谱构造流程

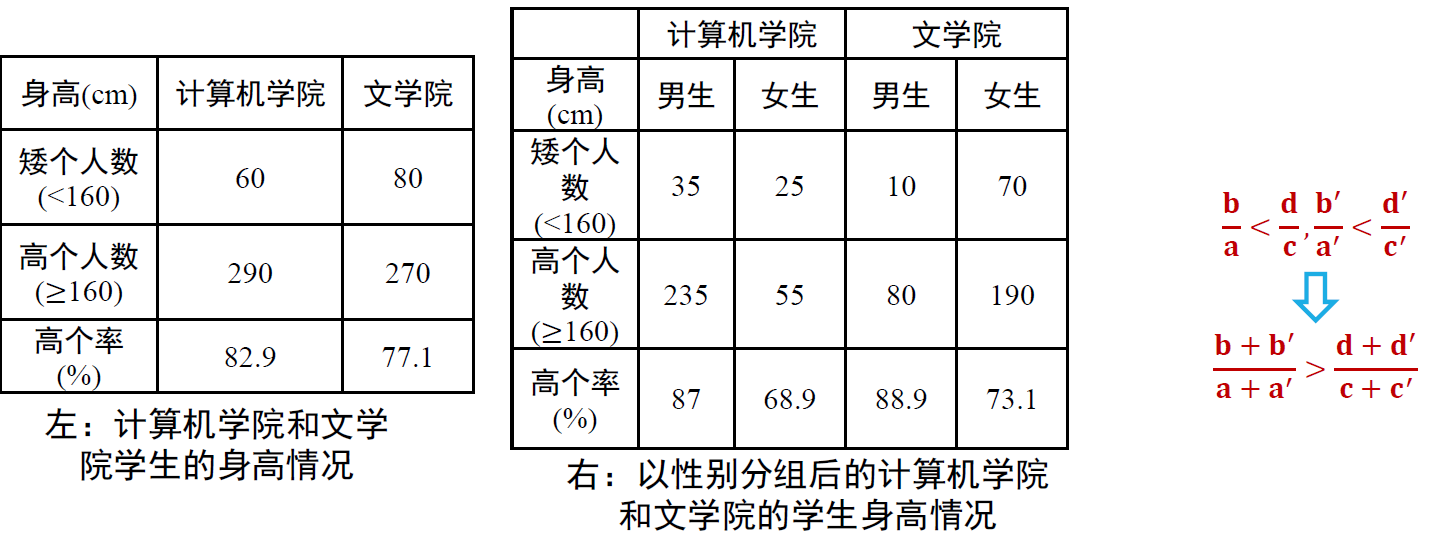


**属性识别与填充**





4. Simpson’s Paradox (辛普森悖论)



🟎计算机学院学生的高个率高于文学院（左表）。

🟎分别比较两所学院男生和女生身高时，却发现计算机学院男生和女生的高个率均低于文学院

🟎在总体样本上成立的某种关系却在分组样本里恰好相反。

🟎右表体现了男生比女生个子高这一现象，如计算机学院和文学院男生高个率都比女生高个率要大。

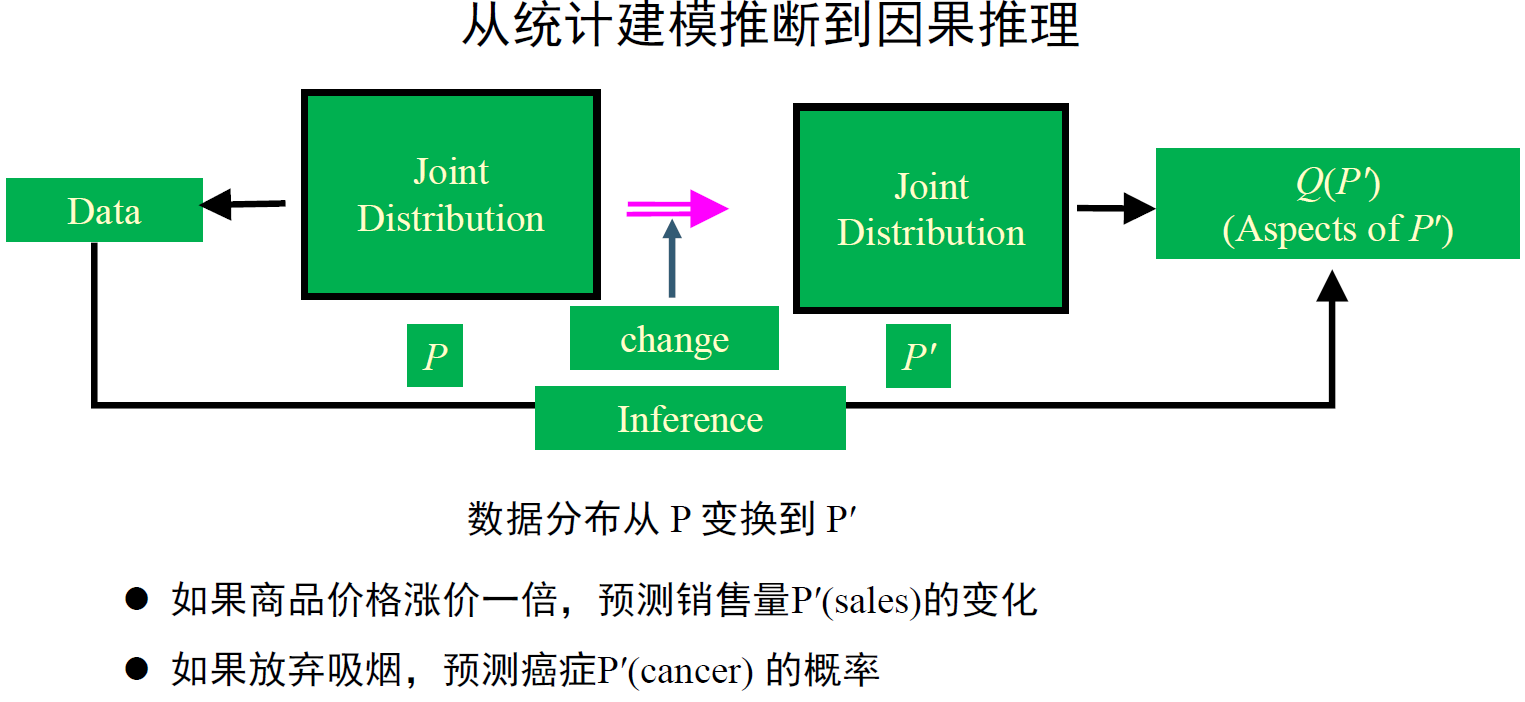
🟎性别会影响专业选择，计算机学院招收的男生多，而文学院招收的女生多。因此，当计算机学院的样本中包含更多的男生，文学院的样本中包含更多的女生，就会看到左表所呈现的情况：计算机学院的高个率高于文学院。

辛普森悖论表明，在某些情况下，忽略潜在的“第三个变量”（如性别就是专业和身高之外的第三个变量），可能会改变已有的结论，而我们常常却一无所知。

从观测结果中寻找引发结果的原因，由果溯因，就是本节要介绍的因果推理

不能只满足于数字或图表，必须考虑数据生成过程----因果模型

5. 因果推理



结构因果模型(structural causal model, SCM):

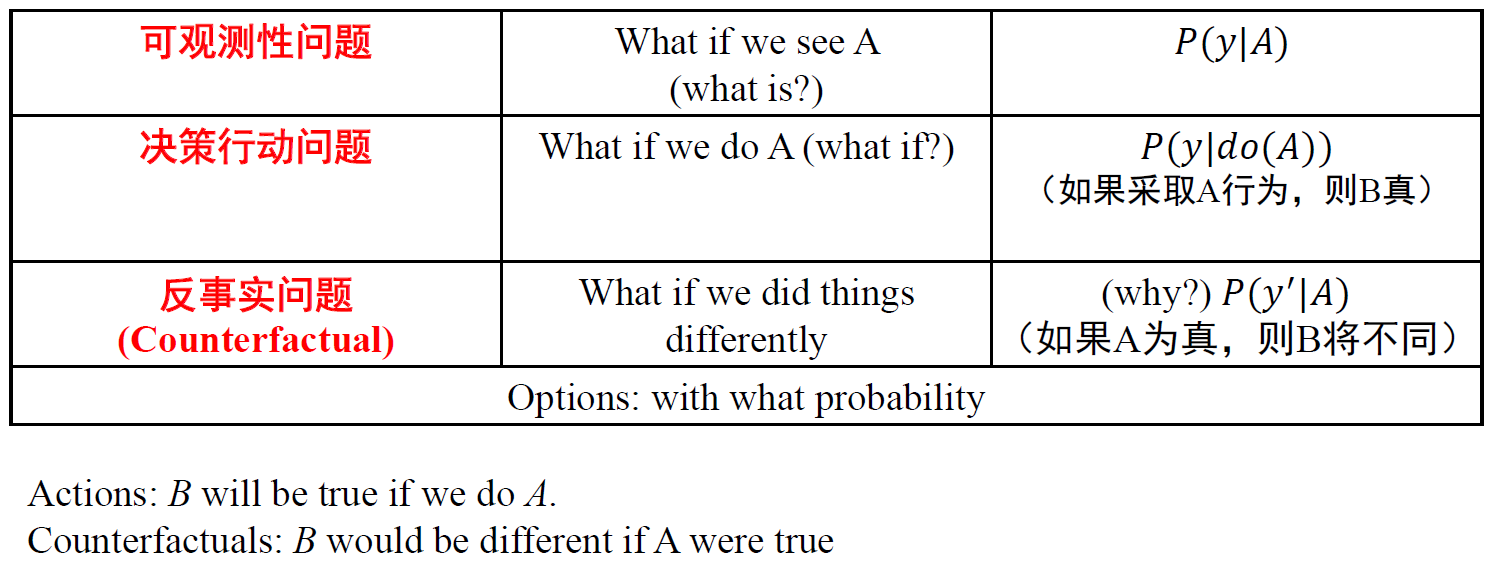
也被称为因果模型（causal model）或Neyman–Rubin因果模型。这一模型最早可追溯于Jerzy Neyman在1923年用波兰语所撰写的博士论文中提出的“潜在结果”（potential outcome）的概念。之后，Donald Rubin发展了“潜在结果”这一概念，并将其和缺失数据的理论联系在一起。

因果图(causal diagram):

由Judea Pearl于1995年提出

每个结构因果模型𝑀都与一个因果图𝐺相对应

6. 因果推理的层级

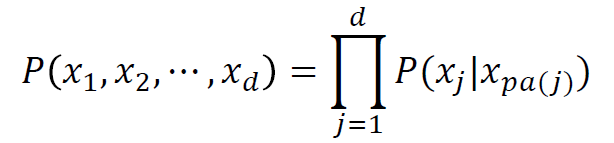


7. 有向无环图(directed acyclic graphs, DAG)

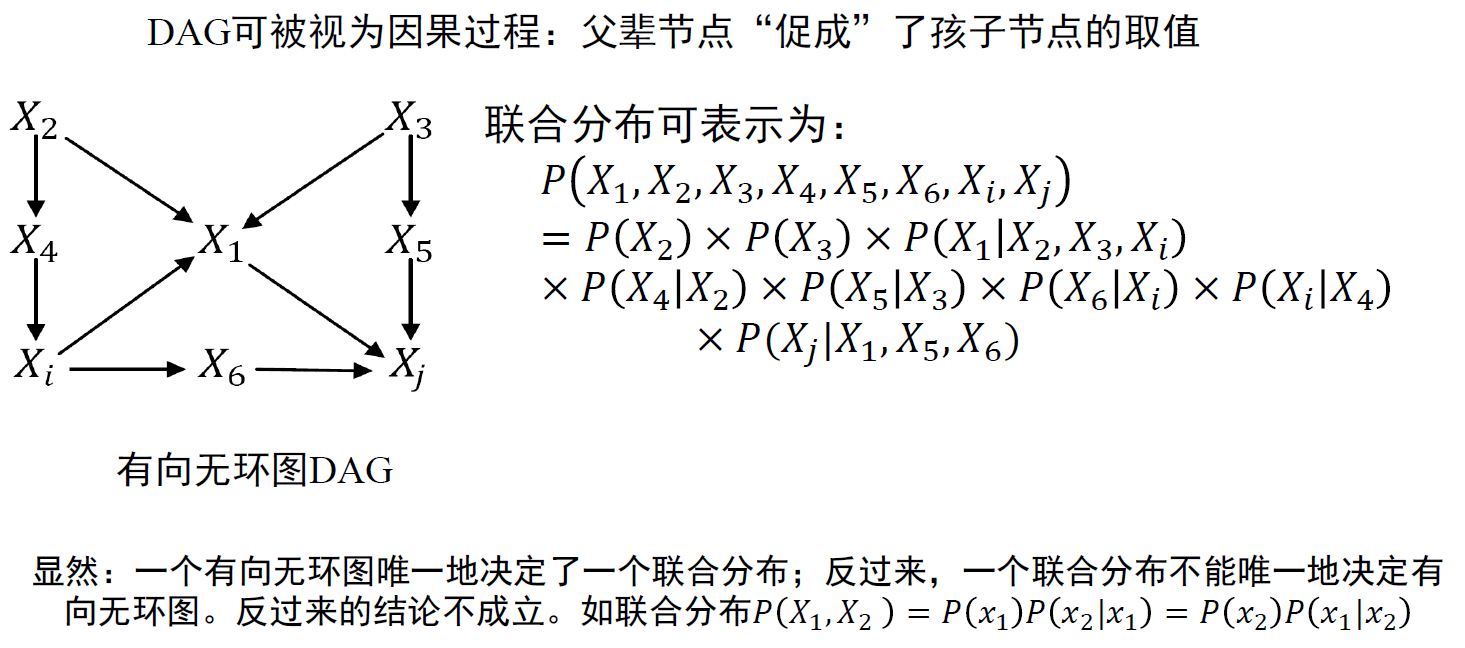
有向无环图指的是一个无回路的有向图，即从图中任意一个节点出发经过任意条边，均无法回到该节点。有向无环图刻画了图中所有节点之间的依赖关系。

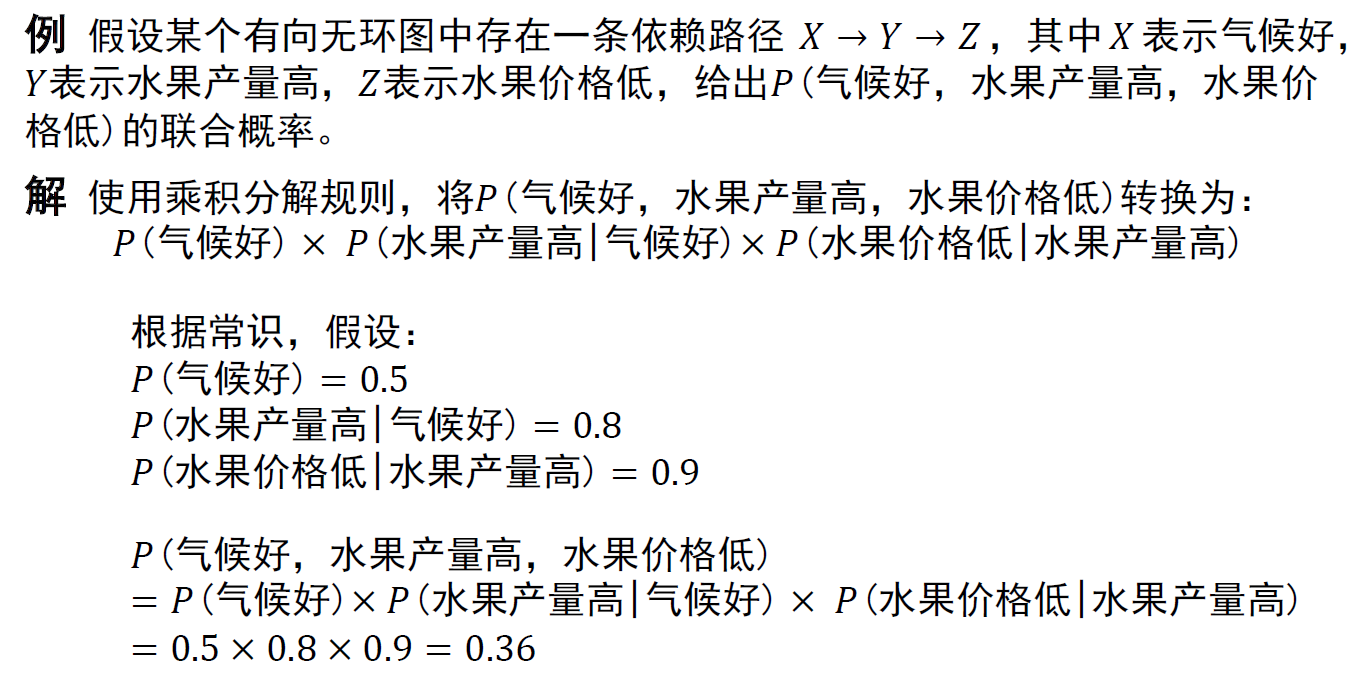
DAG 可用于描述数据的生成机制。这样描述变量联合分布或者数据生成机制的模型，被称为 “贝叶斯网络”（Bayesian network）

对于任意的有向无环图模型，模型中𝑑个变量的联合概率分布由每个节点与其父节点之间条件概率𝑃(𝑐ℎ𝑖𝑙𝑑|𝑝𝑎𝑟𝑒𝑛𝑡𝑠)的乘积给出：



其中，𝑥𝑝𝑎(𝑗)表示节点𝑥𝑗的父节点集合（所有指向𝑥𝑗的节点）

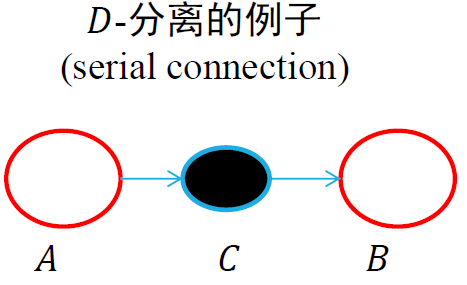




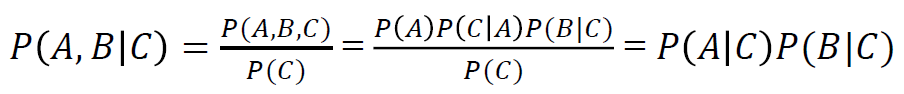
8. 𝐷-分离(directional separation, d-separation)

𝐷-分离用于判断集合𝐴中变量是否与集合𝐵中变量相互独立（给定集合𝐶），记为：𝐴⊥𝐵 | 𝐶

序列连接



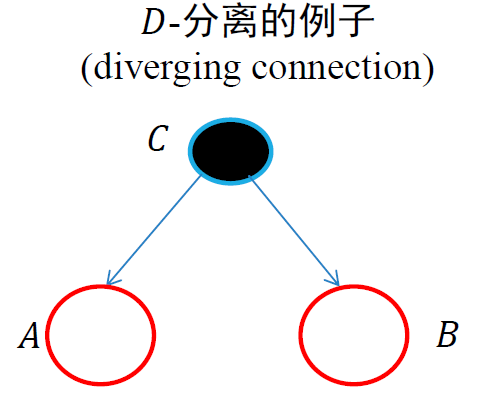
当C取值固定(可观测, observed)，有：



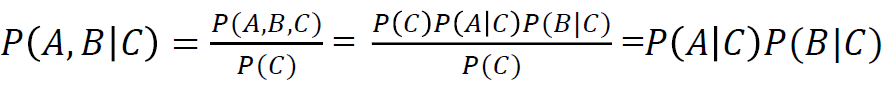
可见A和B在C取值固定情况下，是条件独立的。

注：上式利用了𝑃(𝐴)𝑃(𝐶|𝐴)=𝑃(𝐶)𝑃(𝐴|𝐶)

分叉连接



当C取值固定(可观测, observed)，有：

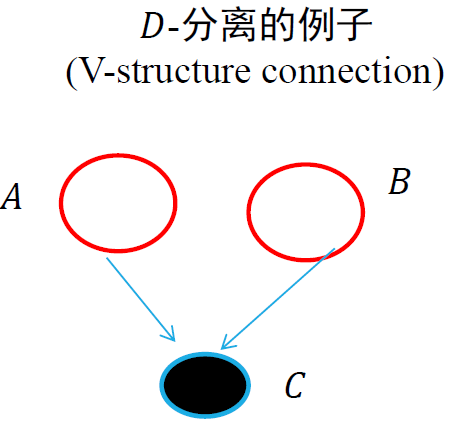


可见A和B在C取值固定情况下，是条件独立的。

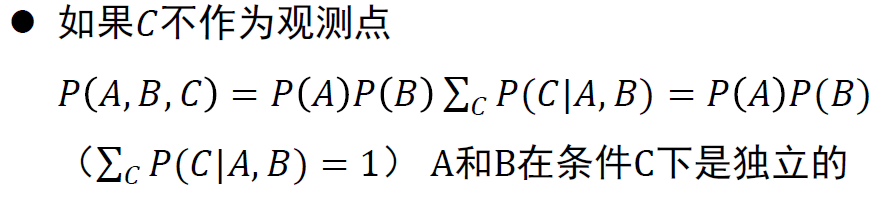
如果C不固定，则有𝑃(𝐴,𝐵)= 𝑃(𝐴𝐵|)𝑃(𝐵|𝐶)𝑃(𝐶)

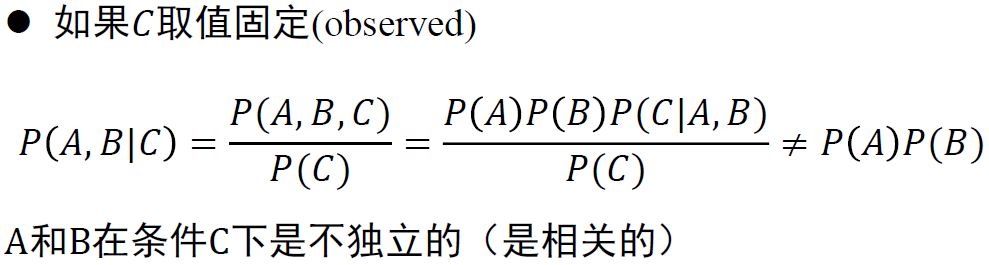
可见，由于𝑃(𝐴,𝐵)≠𝑃(𝐴)𝑃(𝐵)，因此A和B在条件C下不独立的。

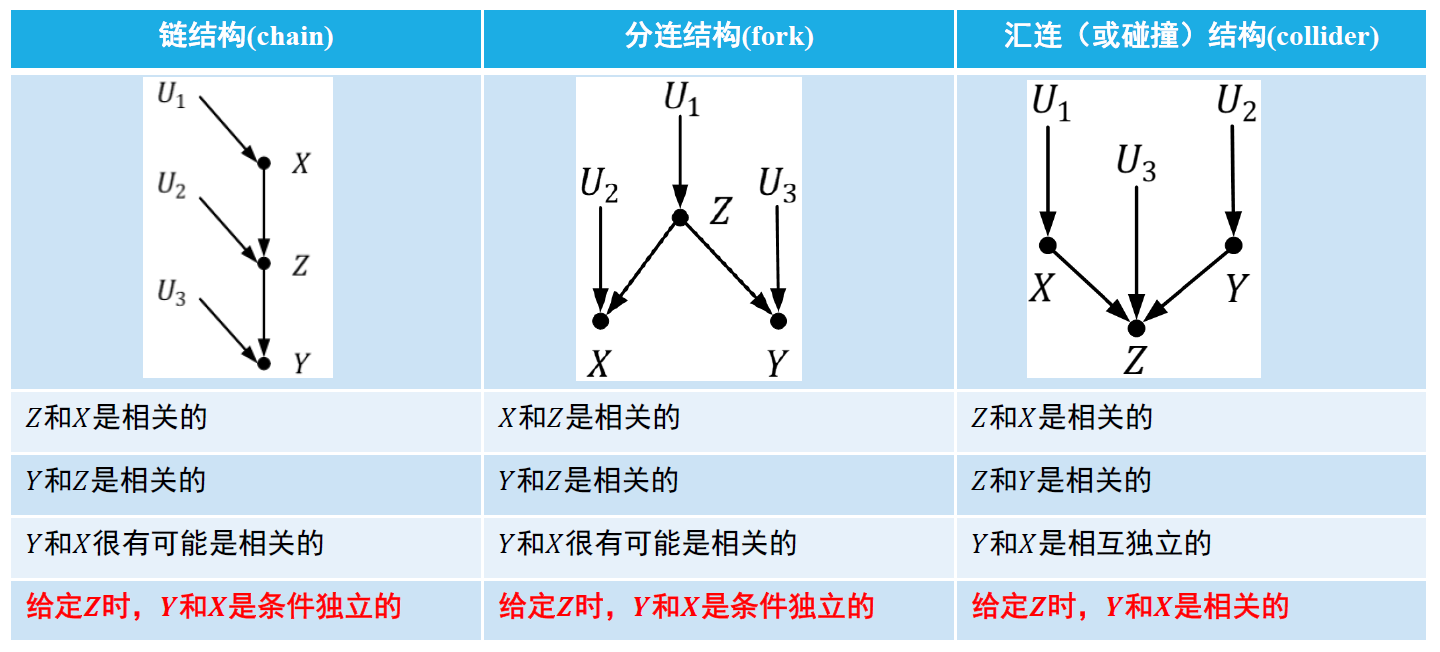
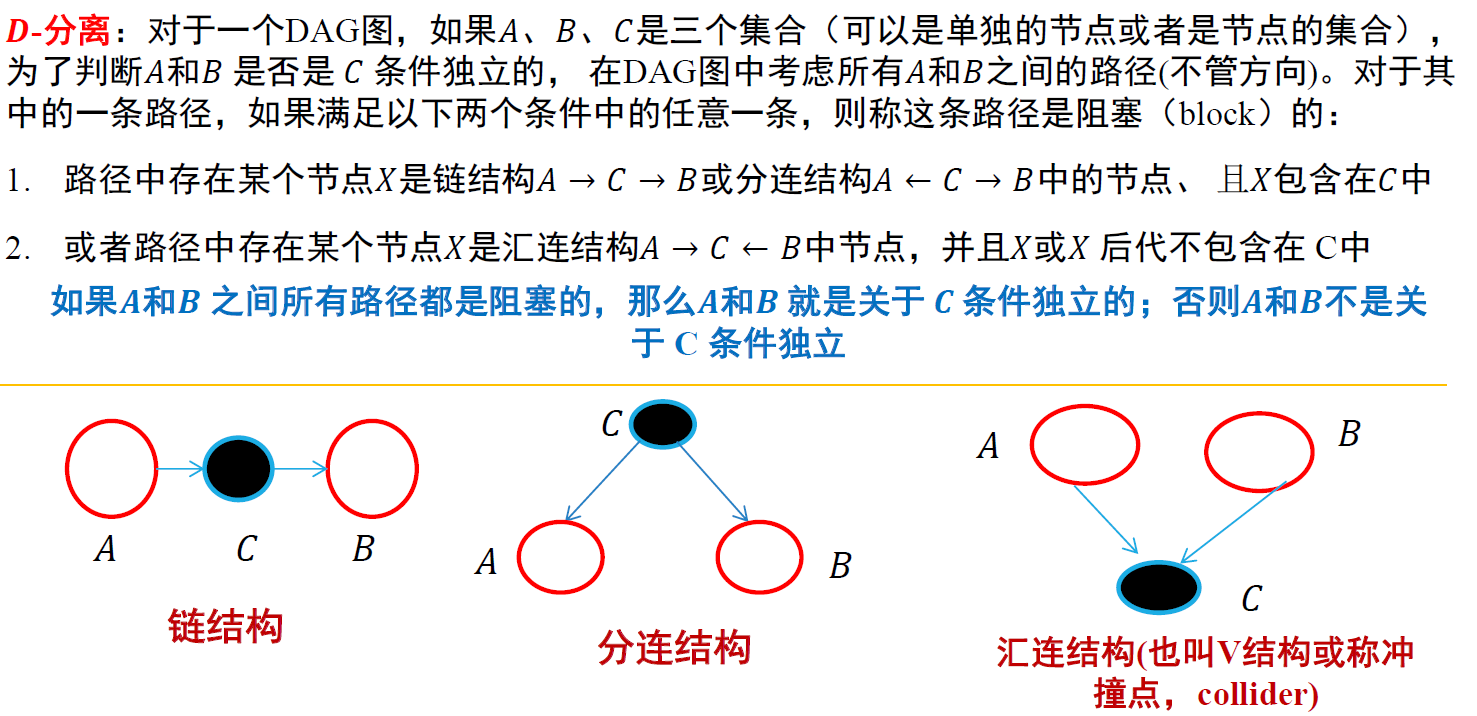
V-结构 连接









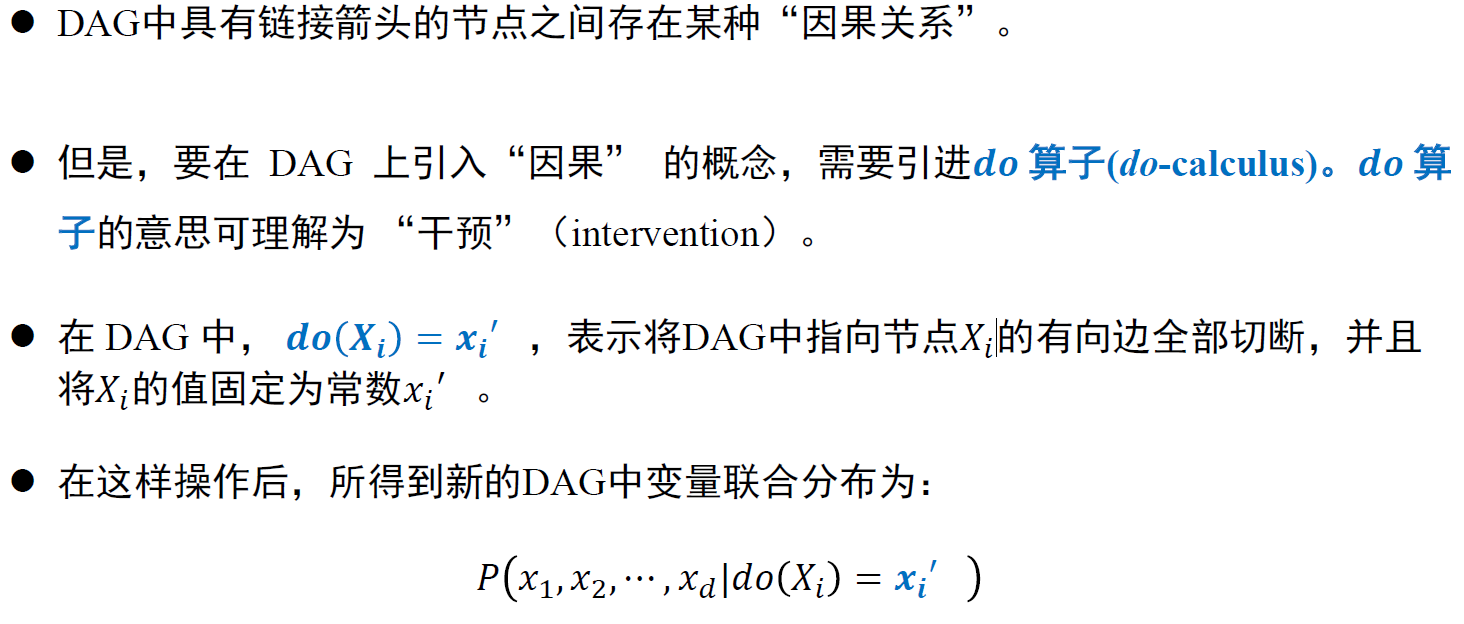


🟎𝐷-分离(directional separation, d-separation)方法可用于判断因果图上任意变量间相关性和独立性。

🟎在因果图上，若两个节点间存在一条路径将这两个节点连通，则称之为是有向连接(d-connected)的；若两个节点不是有向连接的，则称之为是有向分离(d-separated)的，即不存在这样的路径将这两个节点连通。当两个节点是有向分离时，意味着这两个节点相互独立。

🟎若节点𝑋和节点𝑌之间的每一条路径都是阻塞的(blocked)，称节点𝑋和节点𝑌是有向分离的；反之，若存在一条路径是非阻塞的(unblocked)，称节点𝑋和节点𝑌是有向连接的。

9. 干预(intervention)和do算子(do-calculus)



10. 因果效应和因果效应差

