因果推断与大语言模型

李英杰

中国科学院信息工程研究所
INSTITUTE OF INFORMATION ENGINEERING,CAS

2023.12.22

目录

- 因果推断基础
- Causal Inference for LLMs
- LLMs for Causal Inference
- 总结
- 参考文献

贝叶斯公式

贝叶斯公式：

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

$$P(A|B) = P(B|A) * P(A) / P(B)$$

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

$$= \text{似然比} * \text{先验概率}$$

全概率公式：

$$P(B) = \sum_j P(B|A) * A_j$$

贝叶斯公式

贝叶斯公式： $P(D|T) = \frac{P(T|D)}{P(T)} * P(D)$ =似然比 * 先验概率

假设一位四十岁的女性接受 X 光检查以检查是否患有乳腺癌，结果呈阳性。假设 D 患有癌症，证据 T是 X 光检查的结果。

检查灵敏度：$P(T|D) = 73\%$

先验概率：$P(D) = 1/700$

P(T) 是 P(T | D)（患病者中检测呈阳性的概率）和 P(T | ~D)（非患者中检测呈阳性的概率，误报率）的加权平均值。40 岁女性的假阳性率约为 12%

$P(T) = \sum_j P(T|D) * D_j = (73\%) \times (1/700) + (12\%) \times (699/700) \approx 12.1\%$

则得到：$P(D|T) = \frac{73\%}{12.1\%} * (1/700) \approx 1/116$

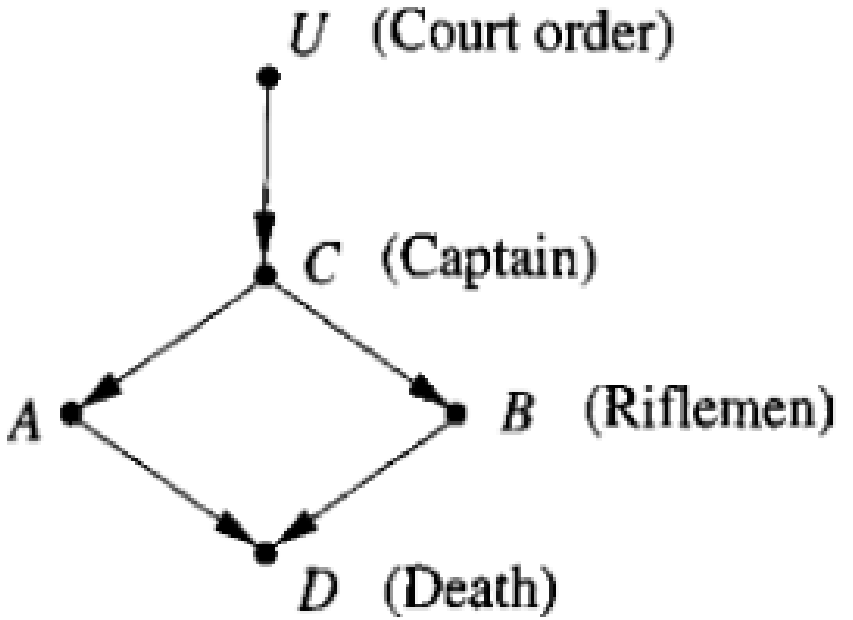
最简单的贝叶斯网络图：D→T

从因果角度看， 前向概率P(T|D)是已知患病后检查阳性的概率，为顺因果方向；而P(D|T) 是通过X光检查的结果反推患病的概率，为逆因果方向（仅为关联路径）。

贝叶斯网络与因果图

共同点：
节点（概率）仅由parents决定；关联流动方式相同——即对于相关性，无论是否顺着箭头方向都可以计算。
不同点：
贝叶斯网络是简化的概率表，单纯表示统计规则，其箭头方向不具有因果意义；
因果图的箭头方向代表原因→结果的因果概念。

因果方向的意义：
决策（因果效应估计）：如（戴眼镜←学习时间长→成绩好）不能通过戴眼镜来提升成绩。
归因（因果发现）：求职被录/拒的原因。户口？性别？学历？
因果量识别、反事实推断、策略学习……



二人行刑队因果图
U：法庭判决是否死亡（未知变量）
C：指挥官是否下达开枪命令
A：士兵A是否开枪
B：士兵B是否开枪
D：死刑犯是否死亡

经典悖论

MONTY HALL PROBLEM（三门问题）

假设一个游戏节目，有三扇门1、2、3。一扇门后面是汽车，其他两扇门后面是山羊。你选择一扇门，比如1号。主持人知道另外两扇门2和3后是否有汽车，然后会打开一扇没有车的门（如果两扇门都没有车，则会在这两扇中随机打开一扇）。假设主持人打开了三号门（门后是羊），这时改变你选择的门（从1改选2）对你有利吗?

思考:

1、主持人的行为对我的初始选择有有影响吗?

2、主持人的行为对车在哪个门有影响吗?

3、改选2号门和保持1号门两个决策获得大奖汽车的概率分别是多少?

经典悖论

MONTY HALL PROBLEM（三门问题）

假设一个游戏节目，有三扇门1、2、3。一扇门后面是汽车，其他两扇门后面是山羊。你选择一扇门，比如1号。主持人知道另外两扇门2和3后是否有汽车，然后会打开一扇没有车的门（如果两扇门都没有车，则会在这两扇中随机打开一扇）。假设主持人打开了三号门（门后是羊），这时改变你选择的门（从1改选2）对你有利吗?

令变量X为初始选择的门，变量Y为后面有汽车的门，变量Z为主持人打开的门。则X,Y,Z取值范围是{1,2,3}，而当前情况是X=1，$Y \in \{1,2,3\}$，Z=3。

解法1：

最直观的方法，因为汽车位置是随机的，所以车在1号门后的概率是1/3，车在2、3号门中的概率是2/3，而主持人排除了2、3号门中一扇没有车的3号，所以2、3号门的整体概率都落在了2号门，即P(Y=2|X=1,Z=3)=2/3，即应该换为2号门。

经典悖论

MONTY HALL PROBLEM（三门问题）

令变量X为初始选择的门，变量Y为后面有汽车的门，变量Z为主持人打开的门。则X,Y,Z取值范围是{1,2,3}，而当前情况是X=1，Y∈{1,2,3}，Z=3。

解法2：排除错误答案法

外推解法1，假设改变游戏规则，有1000扇门，初始仍选1号门，主持人会打开余下999扇门中998扇没有汽车的门（如果999扇门全部没汽车，则随机选取998扇），假设打开了3-1000号门，只留下2号门没打开。

与解法1类似，因为汽车位置是随机的，所以车在1号门后的概率是1/1000，车在其他门中的概率是999/1000，而主持人排除了其他门中所有没汽车的门，所以其他门的整体概率都落在了2号门，即P(Y=2|X=1,Z=3)=999/1000，即应该换为2号门。

经典悖论

MONTY HALL PROBLEM（三门问题）

令变量X为初始选择的门，变量Y为后面有汽车的门，变量Z为主持人打开的门。则X,Y,Z取值范围是{1,2,3}，而当前情况是X=1，Y∈{1,2,3}，Z=3。

解法3：贝叶斯公式

根据已知条件，在主持人打开3号门之后，汽车在1、2号门的概率分别表示为：P(Y=1|X=1,Z=3)和P(Y=2|X=1,Z=3)，分别用贝叶斯公式计算：

$$P(Y=1|X=1,Z=3)=P(Y=1|Z=3)$$ （Y与X无关，车在哪个门后与初始选择无关）

$$=\frac{P(Z=3|Y=1)\,P(Y=1)}{P(Z=3)}$$ （贝叶斯公式）

$$=\frac{1/2*1/3}{P(Z=3)}$$ （注意这个1/2，因为这时3号门是随机选择的）

$$P(Y=2|X=1,Z=3)=\frac{P(Z=3|Y=2)\,P(Y=2)}{P(Z=3)}$$ （同理Y=2也与X取值无关）

$$=\frac{1*1/3}{P(Z=3)}$$ （这里的1是因为若Y=2则主持人只能选3号门）

则已知条件下汽车在2号门的概率是1号门的两倍。

经典悖论

MONTY HALL PROBLEM（三门问题）

假设一个游戏节目，有三扇门1、2、3。一扇门后面是汽车，其他两扇门后面是山羊。你选择一扇门，比如1号。主持人知道另外两扇门2和3后是否有汽车，然后会打开一扇没有车的门（如果两扇门都没有车，则会在这两扇中随机打开一扇）。假设主持人打开了三号门（门后是羊），这时改变你选择的门（从1改选2）对你有利吗?

令变量X为初始选择的门，变量Y为后面有汽车的门，变量Z为主持人打开的门。则X,Y,Z取值范围是{1,2,3}，而当前情况是X=1，Y∈{1,2,3}，Z=3。

为什么主持人的行为会对我的决策产生影响?因果图X→Z←Y（碰撞子，collider），最反直觉的因果图基本结合形式。当以Z为条件时X和Y会打开非因果路径，产生相关性。

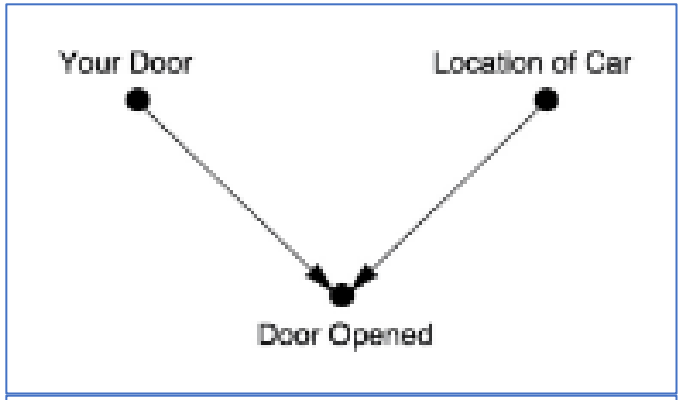


图1

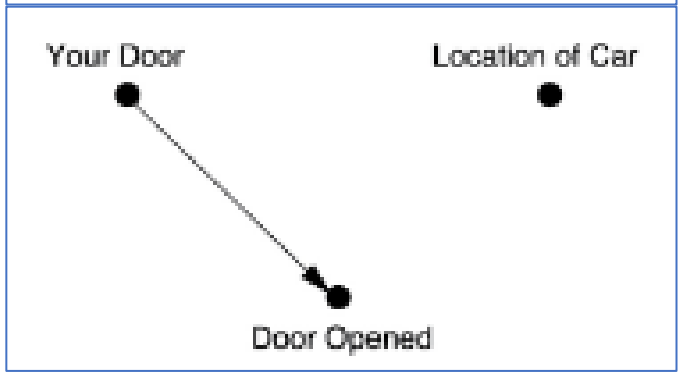


图2

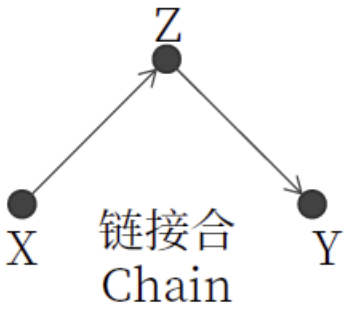
贝叶斯网络的三种基本结构

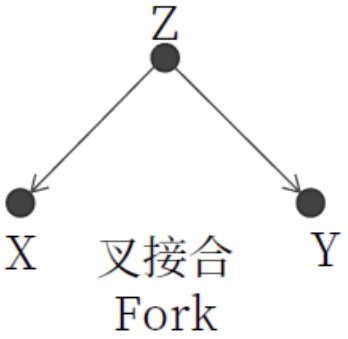
通常研究有向无环图DAG（Directed Acyclic Graph）：

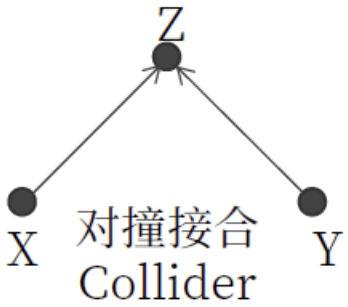
1、X→Z→Y： 链 (Chain) 接合，其中Z被称作“中介变量” (Mediator)

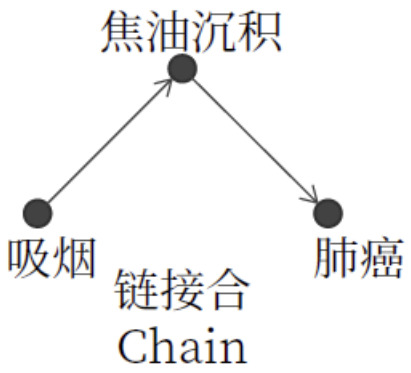
2、X←Z→Y： 叉 (Fork) 接合，其中Z被称作“混杂因子” (Confounder)

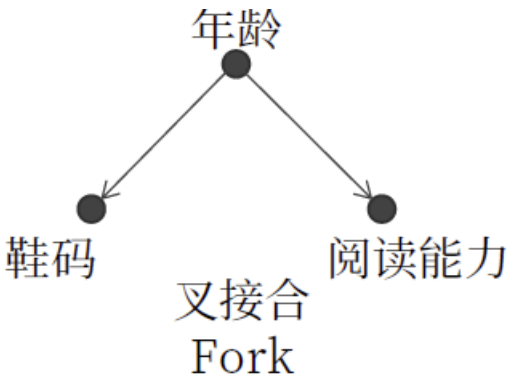
3、X→Z←Y： 对撞 (Collider) 接合，其中Z被称作“对撞因子” (Collider)



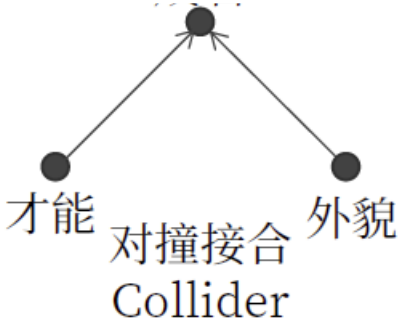








Dating（才华和美貌至少有一个才行）



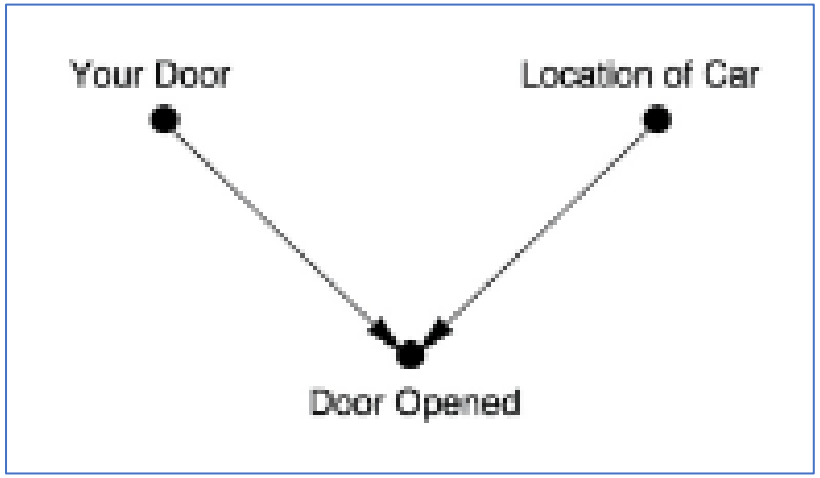
参考《因果推断：曲线拟合之外》2022年——毕冠群

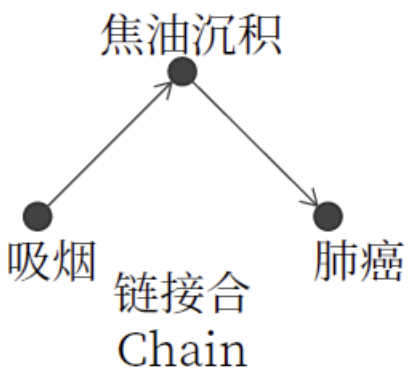
贝叶斯网络的三种基本结构

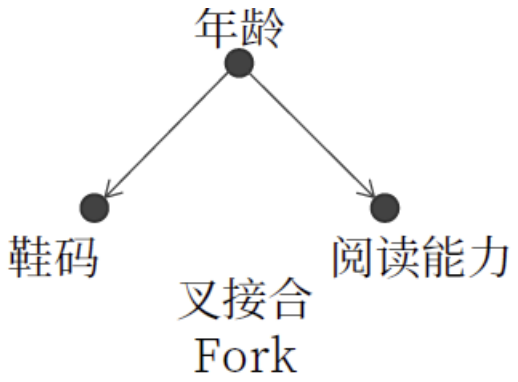
回顾三门问题：令变量X为初始选择的门，变量Y为后面有汽车的门，变量Z为主持人打开的门。则X,Y,Z取值范围是{1,2,3}，而当前情况是X=1，Y∈{1,2,3}，Z=3。

根据已知条件，在主持人打开3号门之后，汽车在1、2号门的概率分别表示为：
P(Y=1|X=1,Z=3)和P(Y=2|X=1,Z=3)

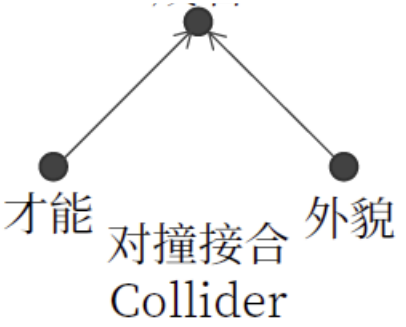
在对撞结合X→Z←Y下以Z为条件，因此X和Y之间产生了相关性。







Dating（才华和美貌至少有一个才行）



d-分离准则

d-分离准则用来判定从X到Y的一条路径是否被变量集Z阻断，这里的路径是包括非因果路径的，即箭头方向不必全是X→⋯→Y

阻断条件（**满足其一则该路径被阻断**）：

1、Z中包含叉结合或链结合

2、Z中不含对撞结合且不含对撞子的后继子孙节点

d-分离准则：如果节点集阻断了两个节点之间的所有路径（包括非因果路径），那么我们则说这两个节点被节点集d-分离了。

非因果路径：X到Y中存在指向X的箭头的路径（即X⋯⋯ ← ⋯⋯ Y）（后门路径）

如果通过d-分离将X和Y之间的非因果路径全部阻断，则可以得到X和Y之间真正的因果关系，这时直接可以得到因果关系。

A和J之间的这条路径是否被阻断，如何阻断？

$$A \leftarrow B \leftarrow C \rightarrow D \leftarrow E \rightarrow F \rightarrow G \leftarrow H \rightarrow I \rightarrow J?$$

d-分离准则

d-分离准则用来判定从X到Y的一条路径是否被变量集Z阻断，这里的路径是包括非因果路径的，即箭头方向不必全是X→⋯→Y

阻断条件（满足其一则该路径被阻断）：

1、Z中包含叉结合或链结合

2、Z中不含对撞结合且不含对撞子的后继子孙节点

d-分离准则：如果节点集阻断了两个节点之间的所有路径（包括非因果路径），那么我们则说这两个节点被节点集d-分离了。

非因果路径：X到Y中存在指向X的箭头的路径（即X⋯⋯ ← ⋯⋯ Y）（后门路径）

如果通过d-分离将X和Y之间的非因果路径全部阻断，则可得到X和Y之间真正的因果关系，这时相关性→因果关系。

$$A \leftarrow B \leftarrow C \rightarrow D \leftarrow E \rightarrow F \rightarrow G \leftarrow H \rightarrow I \rightarrow J?$$

其实A和J之间是自然阻断的，只要出现一个结合处阻断则路径关闭，可以认为Z是空集，则满足条件2（D处和G处的对撞结合阻断了路径）。若Z包含C和G，则只要Z中再包含B或C或E或F或H或I中任意一个即可阻断。

打开非因果路径的情况：Z={D,G}

d-分离准则

判断X和Y直接是否有非因果路径，如果有，如何阻断？

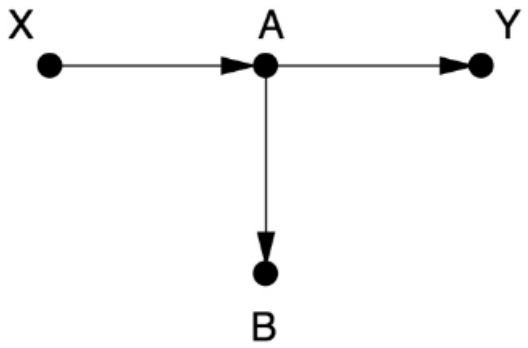


图1

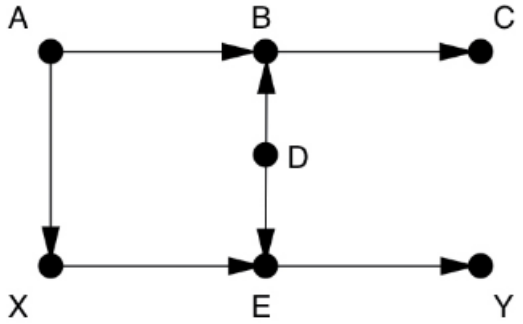


图2

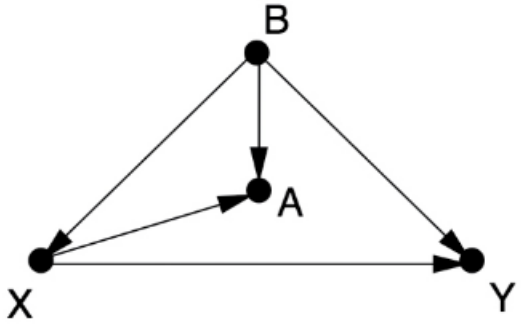


图3

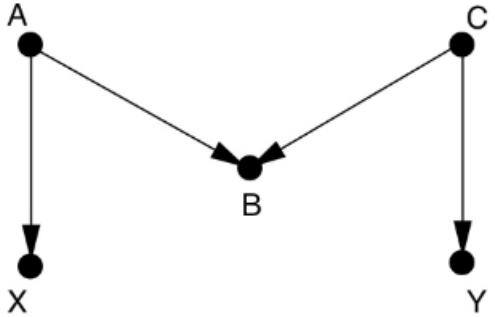


图4
（M型偏差）

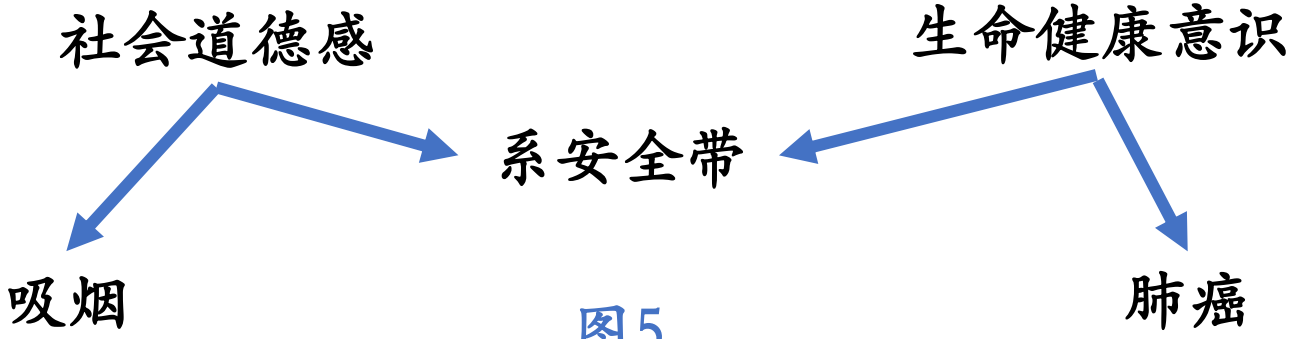


图5
（M型偏差）

经典悖论2

SIMPSON’S PARADOX（辛普森悖论）

“BBG”悖论bad and bad but good？根据是否服药分为实验组treatment和对照组control(不服药)

	Control Group (No Drug)		Treatment Group (Took Drug)	
	Heart attack	*No heart attack*	*Heart attack*	*No heart attack*
Female	1	19	3	37
Male	12	28	8	12
Total	13	47	11	49

对照组中有 5%(1/20)的女性心脏病发作，而服药的女性中病发概率为 7.5%(3/40)，因此，该药物会引发女性心脏病发(bad)；在男性中，对照组中有 30% (12/40)病发，而实验组中有 40% (8/20)的人病发，因此，该药物会引发男性心脏病发(bad) 。

但是，从第三行的总体结果看：在对照组中，21.7%(13/60) 的人病发，但在实验组中只有 18.3% (11/60)的人病发。因此，如果用总体情况判断，药物 D 似乎可以降低整体心脏病发作的风险(good)。 Bad and Bad but Good？总体or局部？

经典悖论2

Simpson’s reversal （辛普森逆转） ≠ Simpson’s paradox （辛普森悖论）

辛普森逆转：错误的不等式

如果 A/B > a/b 且 C/D > c/d，则 (A + C)/(B + D) > (a + c)/(b + d) **不成立！**

病发概率	对照组	服药组
女	1/20	3/40
男	12/40	8/20
总体	13/60	11/60

$1/20<3/40$

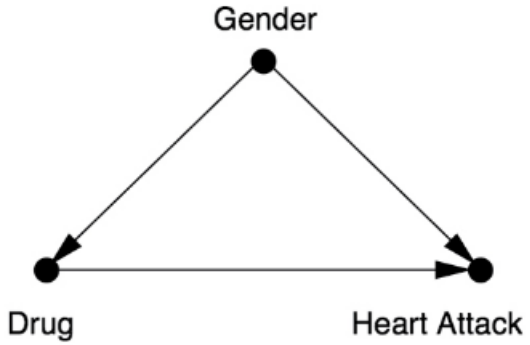
$12/40<8/20$

但是：$\frac{1+12}{20+40} > \frac{3+8}{40+20}$

经典悖论2

Simpson’s reversal （辛普森逆转） ≠ Simpson’s paradox （辛普森悖论）

辛普森悖论：不是辛普森逆转导致，而是虚假的因果关系导致的。



辛普森悖论的因果图

病发概率	对照组	服药组
女	1/20	3/40
男	12/40	8/20
总体	13/60	11/60

根据我们前面了解的d-分离准则，性别处的叉结合（混杂因子）导致了药物←性别→心脏病的非因果路径（不同性别的服药比例不同），因此干扰了我们对于药物与心脏病发之间因果关系的判断。根据d-分类准则，令Z={Gender}即可。

后门准则

Backdoor Criterion（后门准则）

给定一个有向无环图G，以及G中的一对有序变量X和Y，如果一组变量Z满足以Z为条件阻断所有X和Y之间的后门路径，那么变量Z满足关于(X,Y)的后门准则。

Backdoor Adjustment（后门调整公式）

如果变量Z满足关于(X,Y)的后门准则，那么X对Y的因果关系可以写作公式：

$$P(Y=y|do(X=x))=\sum_z P(Y=y|X=x,Z=z)P(Z=z)$$

Do运算：删除指向X的所有箭头

原始因果图　　Do运算

经典悖论2

病发概率 Y=1	对照组 X=0	服药组 X=1
女	1/20	3/40
男	12/40	8/20
总体	13/60	11/60

应用后门调整公式计算辛普森悖论

根据因果图可知混杂因子为性别Z=Gender，假设男女比例=1:1

则 $P(Y=1|do(X=1))$ ——实验组（服药）的因果效应

$=\sum_z P(Y=y|X=x,Z=z)P(Z)$

$=P(Y=1|X=1,Z=女)P(Z=女)+P(Y=1|X=1,Z=男)P(Z=男)$

$=3/40*0.5+8/20*0.5$

$=23.75\%$

$P(Y=1|do(X=0))$ ——对照组（不服药）的因果效应

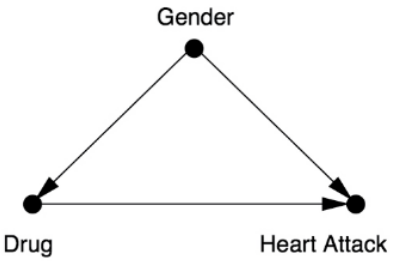
$=\sum_z P(Y=y|X=x,Z=z)P(Z)$

$=P(Y=1|X=0,Z=女)P(Z=女)+P(Y=1|X=0,Z=男)P(Z=男)$

$=1/20*0.5+12/40*0.5$

$=17.5\%$

得到$P(Y=1|do(X=1)) > P(Y=1|do(X=0))$，即从总体上计算，服药也提高了心脏病发的概率，即符合直觉的BAD BAD then BAD



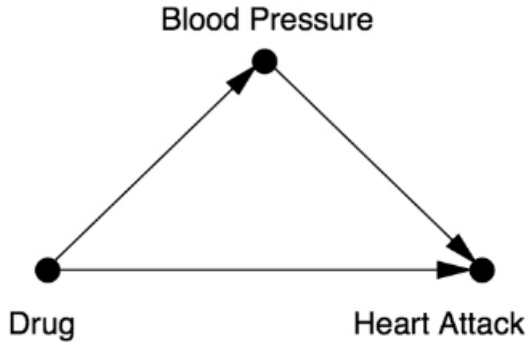
辛普森悖论因果图

经典悖论2

病发概率 Y=1	对照组 X=0	服药组 X=1
女	1/20	3/40
男	12/40	8/20
总体	13/60	11/60

辛普森悖论不等于辛普森逆转

数据表格完全相同，但是对应的变量和因果关系不同。将性别变量换为
介变量）。



辛普森悖论
不同的因果图

	Control Group (No Drug)		Treatment Group (Took Drug)	
	Heart attack	*No heart attack*	*Heart attack*	*No heart attack*
Low blood pressure	1	19	3	37
High blood pressure	12	28	8	12
Total	13	47	11	49

这时我们可以直接利用总体的结果18.3% (11/60) <21.7%(13/60) 来说明实验组（服药）降低了心脏病风险，因为因果图中没有非因果路径。

相同的数据≠相同的结论

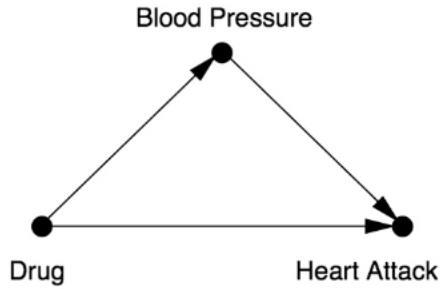
不同的因果关系→不同的数据生成过程

经典悖论2

辛普森悖论不等于辛普森逆转

	Control Group (No Drug)		Treatment Group (Took Drug)	
	Heart attack	*No heart attack*	*Heart attack*	*No heart attack*
Low blood pressure	1	19	3	37
High blood pressure	12	28	8	12
Total	13	47	11	49

病发概率 Y=1	对照组 X=0	服药组 X=1
低血压	5%	7.5%
高血压	30%	40%
总体	21.7%	18.3%



Bad Bad Good? 分别以高血压和低血压为条件看条件概率，则发病率都上升，如果加权计算将得到和性别情况类似的结果（取决于Z=高低血压比例）。但是根据因果图这样计算是错误的（可以理解为只考虑了药物→发病率的直接因果效应），而药物的实际作用是通过降低血压间接降低发病率。

相同的数据≠相同的结论

不同的因果关系→不同的数据生成过程

反事实

反事实问题

我们让 S 代表薪资，EX 代表工作经验(年)，ED 代表教育程度，假设有三个值：0＝高中，1＝大学，2＝研究生。有如表格所示的数据，其中带有下标的S的已知值，代表该员工实际教育程度下的工资，？为反事实问题，例如对于Alice，她本来是高中学历，如果她读了大学ED =1，那么S_1=?

Employee (u)	*EX*(u)	*ED*(u)	S_0(u)	S_1(u)	S_2(u)
Alice	6	0	$81,000	?	?
Bert	9	1	?	$92,500	?
Caroline	9	2	?	?	$97,000
David	8	1	?	$91,000	?
Ernest	12	1	?	$100,000	?
Frances	13	0	$97,000	?	?
etc.					

反事实

反事实问题

Employee (u)	EX(u)	ED(u)	S_0(u)	S_1(u)	S_2(u)
Alice	6	0	\$81,000	?	?
Bert	9	1	?	\$92,500	?

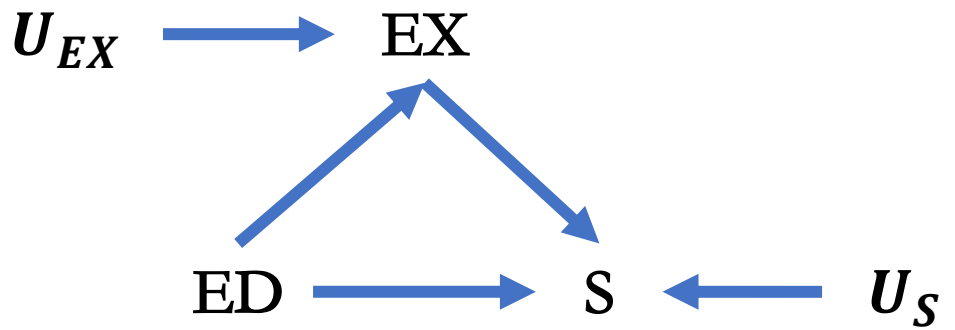
我们让 S 代表薪资，EX 代表工作经验(年)，ED 代表教育程度，假设有三个值：0 = 高中，1 = 大学，2 = 研究生。有如表格所示的数据，其中带有下标的S的已知值，代表该员工实际教育程度下的工资，？为反事实问题，例如对于Alice，她本来是高中学历，如果她读了大学ED =1，那么S_1=?

根据数据进行线性拟合得到近似平面：$S = \$65{,}000 + 2{,}500 \times EX + 5{,}000 \times ED$

那么可以得到S_1 =65000+2500*6+5000*1=85000 ,符合我们的直觉，学历高了工资也高了。

但是我们画出因果图：

从因果图中可以看到未观测的变量U_{EX}可以认为是年龄、学习能力等会影响EX的因素，U_S则可以看作人脉等其他未观测的变量。



根据因果图可以将原本假设的线性方程改写为:

$EX + 4 \times ED = 10 + U_{EX}$

$S = 65000 + 2500 \times EX + 5000 \times ED + U_S$

反事实

Employee (u)	EX(u)	ED(u)	S_0(u)	S_1(u)	S_2(u)
Alice	6	0	$81,000	?	?

反事实计算的步骤：

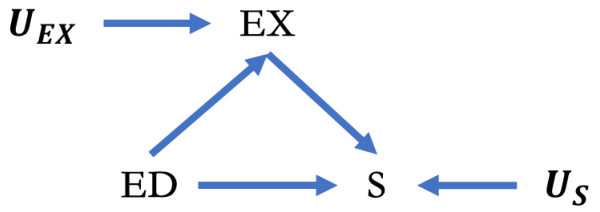
1、外推（Abduction）：使用证据E=e来确定U的值

2、干预（Action）：通过用X=x来替换原来模型M中变量X的表达式，从而修改原模型M为Mx。

3、预测（Prediction）：使用修改后的模型Mx和第一步计算出的U值来计算Y值。

（1）$EX + 4 \times ED = 10 + U_{EX}$

（2）$S = 65000 + 2500 \times EX + 5000 \times ED + U_S$



1、外推：应用已知条件计算Alice的U_{EX} 和U_S

由（1）计算得U_{EX}=-4；由（2）计算得U_S =1000

2、干预：令其拥有学士学位ED(Alice) = 1，并更新因果图，由于没有指向ED的箭头，则Mx=M

3、预测：使用新的因果模型Mx计算S

$EX_{ED=1}$ (Alice) = $10 + U_{EX} - 4 \times ED = 10\text{-}4\text{-}4 \times 1 = 2$

$S = 65000 + 2500 \times EX + 5000 \times ED + U_S$

$= 65000 + 2500 \times 2 + 5000 \times 1 + 1000 = 76000 < 85000$

使用因果图和反事实方法建模后得到的结果更符合常理。

因果之梯

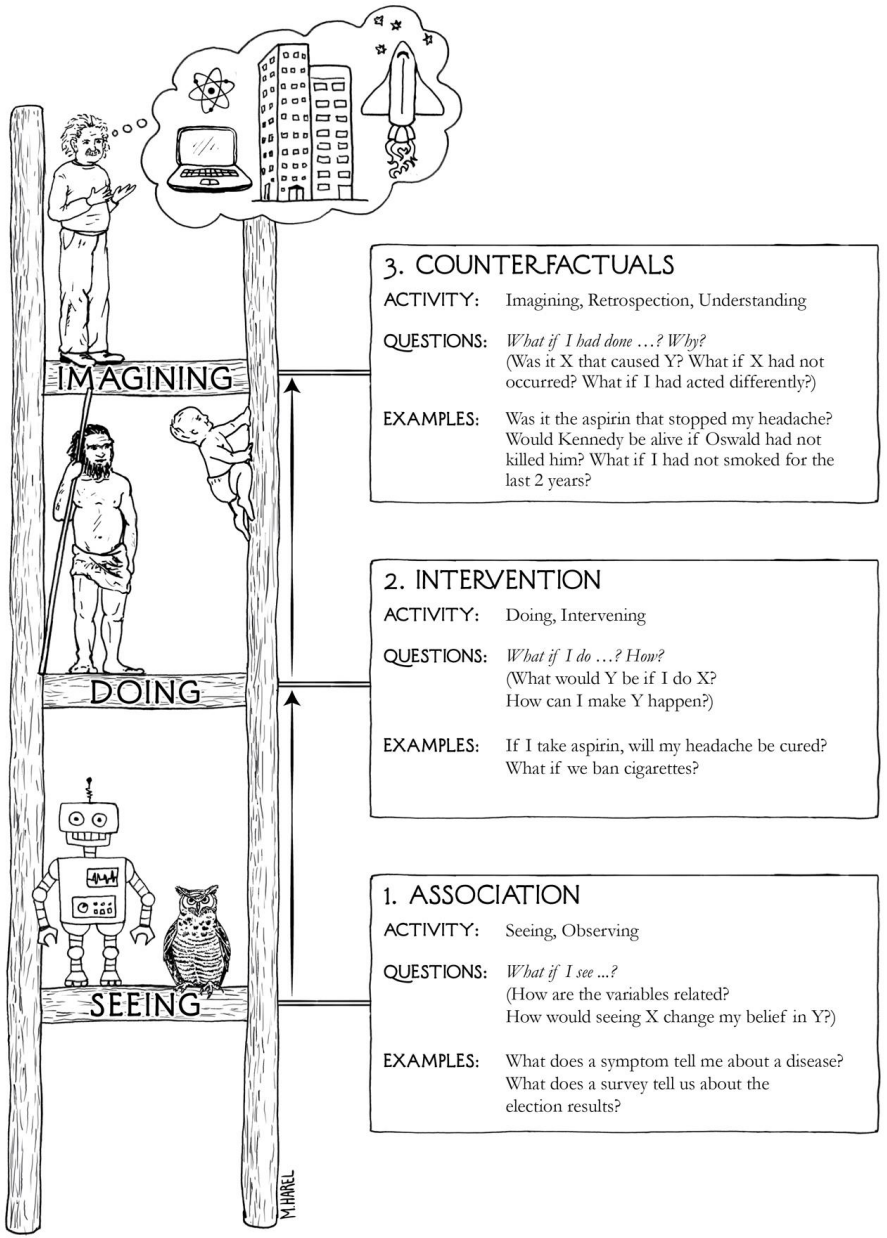
Association→Intervention→Counterfactuals
相关→干预→反事实

Association即仅从数据中发现相关性知识；

Intervention则需要引入因果模型，如结构因果模型等，进行do运算（删去指向X的边）；

Counterfactuals则需要通过已知来估计U（外生变量即未观测到的变量），并通过do-calculus来预测结果。

2层与3层的区别，do运算是总体，反事实则是个体性的…

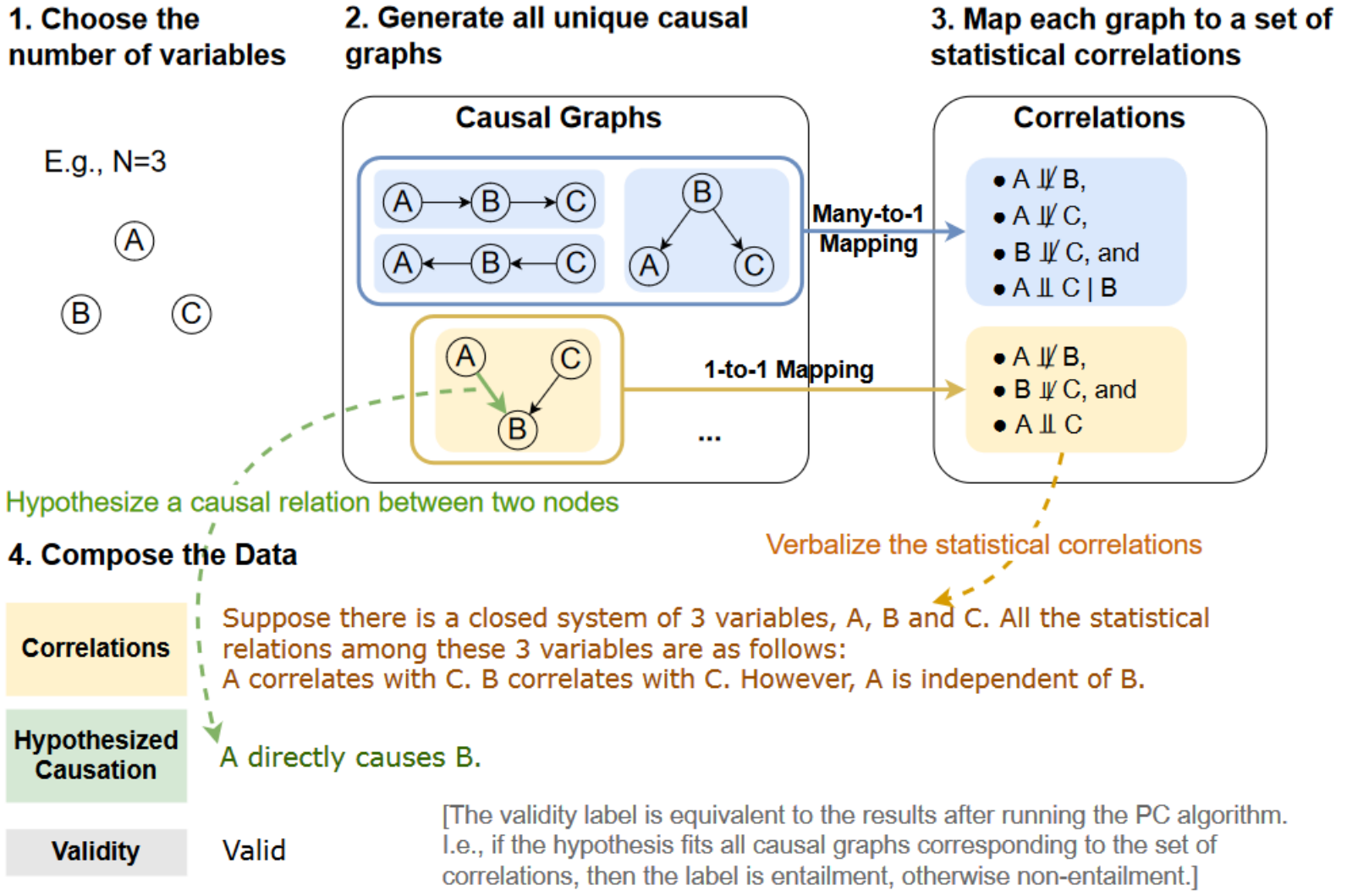


评估大模型的因果推断能力

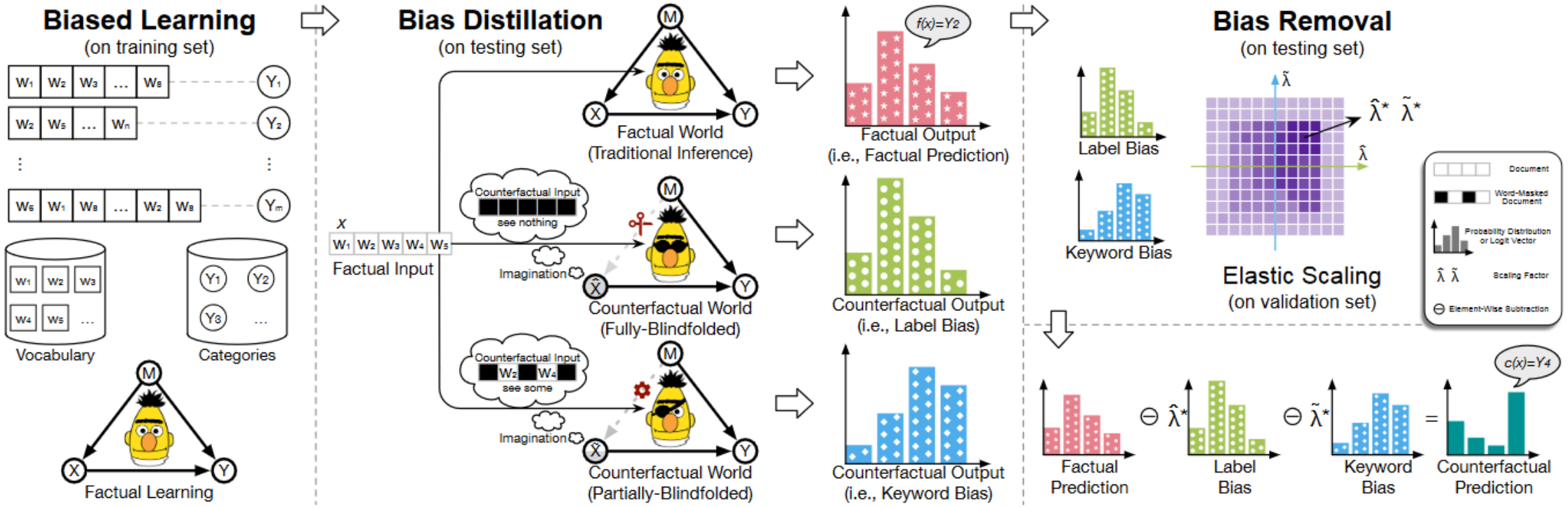
Can Large Language Models Infer Causation from Correlation?

给定总变量节点数量，遍历因果图来得出一些相关性结论生成数据，其中对撞结合可以生成有明确因果关系的条件；

根据生成的数据来给出相关性结论，然后让大模型根据已知的相关性条件来判定给出的因果假设是否正确。



利用反事实数据去偏



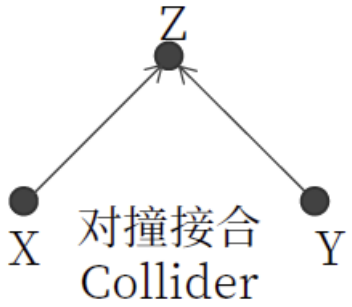
通过对反事实数据——原文全部MASK和MASK部分词来获得数据的bias，再通过处理概率得到去偏之后的数据

因果发现

因果发现方法：

1、基于独立性的方法

可以理解为利用对撞结合若X、Y相互独立，但是X与Z、Y与Z均不独立，则可以判定X、Z、Y形成对撞子，确定X→Z和Y→Z两条边



代表方法：PC算法，不需要穷举独立性

2、加性噪声模型

加性噪声模型：如果X和Y的关系可以用如下的一个函数加一个噪声的SCM形式来表示的话，那么我们称联合概率分布$P_{X,Y}$满足X到Y的加性噪声模型：

$$Y = f_Y(X) + N_Y, N_Y \perp\!\!\!\perp X$$

基于加性噪声模型的可辨识性定义，判断X和Y之间因果关系的方法如下：

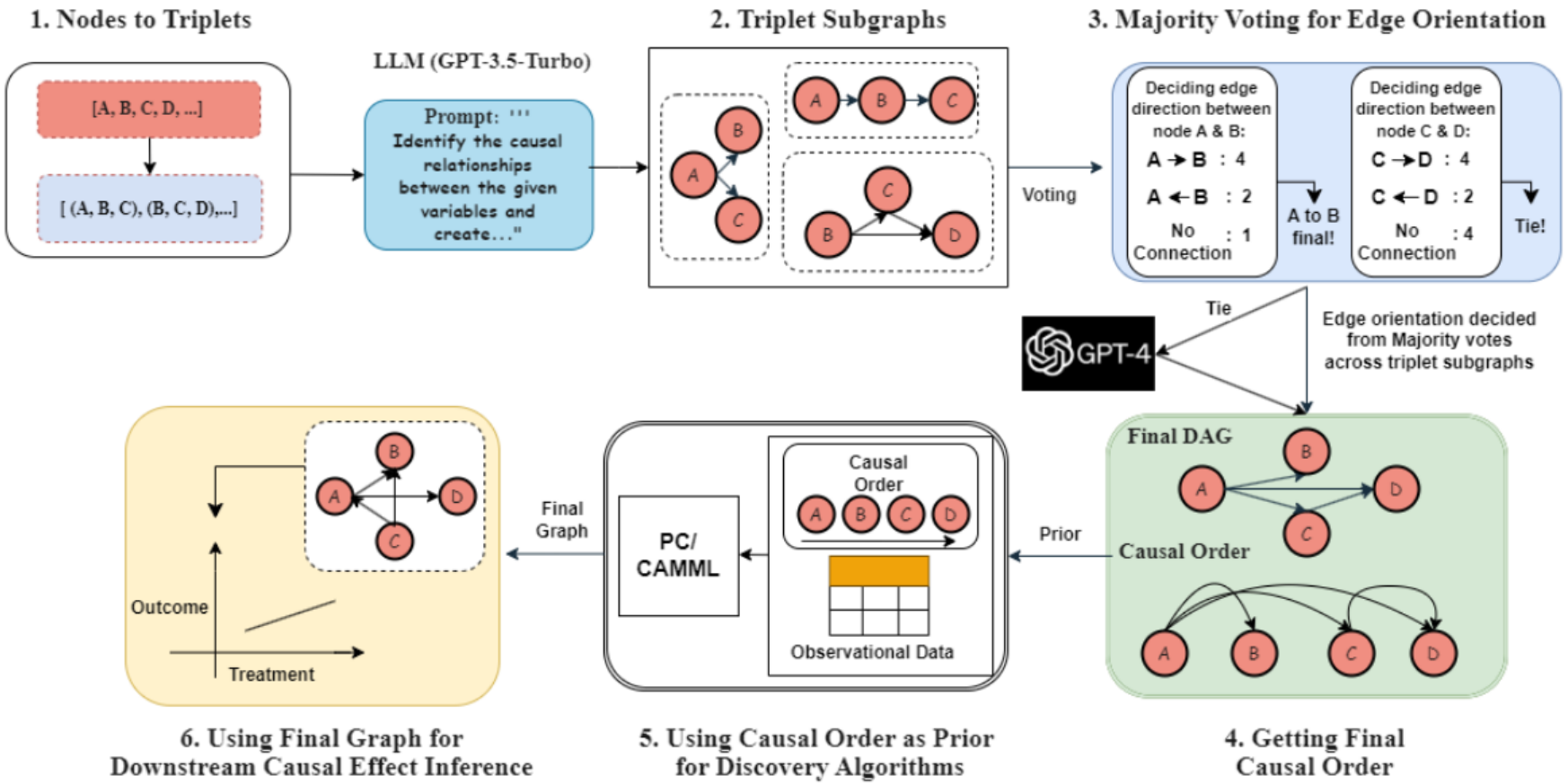
1、在X上回归Y，即用某种回归算法将Y表示为一个关于X的函数f_Y，加上某个噪声。

2、测试Y-f_Y是否与X独立。

3、重复以上步骤但反过来在X上回归Y 。

4、如果这两种情形的测试结果，某一个独立，另一个不独立，那么独立的那个就是因果关系方向。

因果发现



使用大模型通过三个变量的子图来投票确定变量之间的因果顺序，辅助因果发现算法来提升因果发现能力

LLM与因果推断

LLMs for Causal Inference：

利用大模型的知识辅助因果发任务，如构建因果图

生成反事实数据

对已有因果图合理性进行分析，归因分析等

注意：大模型更多是从语义角度根据知识去记忆和推理因果关系，而不是利用数值化的统计数据得到因果关系，但是仍对因果推断有很大的帮助。

Causal Inference for LLMs：

利用反事实数据辅助大模型训练增强其因果推断能力和鲁棒性

根据因果推断方法对大模型结果进行纠偏，增加生成文本的可信度等

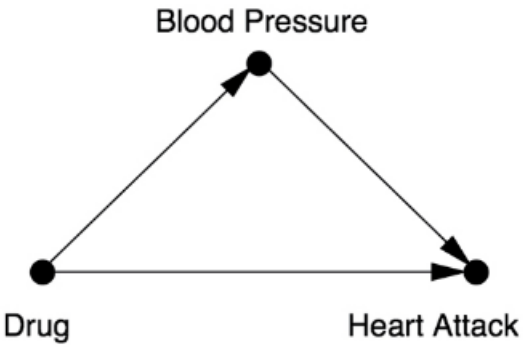
典型的反事实提问：如果珠穆朗玛峰下降1000米，世界最高峰TOP5将会变成什么样?

因果与逻辑

目前的个人理解

逻辑推理是符号和规则，因果推断是从现实世界挖掘因果联系。如通常根据逻辑推理由A→B,B→C，则可以得到A→C。

但是从因果角度看则不同，直接因果和间接因果是完全不同的概念，需要具体看ABC都是什么，是否有统计数据等。因此因果关系是不具备传递性的（但在部分工作及数据集中认为其符合传递性，便于分析）。（附庸的附庸不是附庸，即A→B,B→C的关系中，A不一定可以直接影响C）



逻辑关系中通常条件蕴含结论，如A是正整数则必然A大于0。

因果关系的因只要可以影响果即可，不一定要使果一定发生，充分性和必要性都可以认为是因果关系，如法律判定责任时，A开枪攻击B导致B死亡，充分性强；A乱扔香蕉皮导致B滑倒受伤，在判定A的责任时A行为的充分性不强，但对于B滑倒有必要性。

参考文献

[1] A. Vashishtha, A. G. Reddy, A. Kumar, S. Bachu, V. N. Balasubramanian, and A. Sharma, “Causal Inference Using LLM-Guided Discovery.” arXiv, Oct. 23, 2023. doi: 10.48550/arXiv.2310.15117.
[2] Z. Jin et al., “Can Large Language Models Infer Causation from Correlation?” arXiv, Jun. 09, 2023. doi: 10.48550/arXiv.2306.05836.
[3] E. Kıcıman, R. Ness, A. Sharma, and C. Tan, “Causal Reasoning and Large Language Models: Opening a New Frontier for Causality.” arXiv, May 08, 2023. doi: 10.48550/arXiv.2305.00050.
[4] C. Qian, F. Feng, L. Wen, C. Ma, and P. Xie, “Counterfactual Inference for Text Classification Debiasing,” in Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Online: Association for Computational Linguistics, Aug. 2021, pp. 5434 – 5445. doi: 10.18653/v1/2021.acl-long.422.
[5] J. Pearl, “An Introduction to Causal Inference,” Int J Biostat, vol. 6, no. 2, p. 7, Feb. 2010, doi: 10.2202/1557-4679.1203.
[6] J. Pearl, “Causal inference in statistics: An overview,” Statistics Surveys, vol. 3, no. none, pp. 96 – 146, Jan. 2009, doi: 10.1214/09-SS057.[7] CAUSALITY: Models, Reasoning, and Inference Second Edition. [8] Causal Inference in Statistics. [9] M. Willig and M. Zečević, “Causal Parrots: Large Language Models May Talk Causality But Are Not Causal”.

敬请批评指正！