团队文档

团队活动

团队微博

因果推断简介(一):干预与因果推断 🖫 xuenanwang 2019年08月23日 13:52 浏览(843) 🔘 收藏(27) 🔘 评论(1) 🧠 分享 🔻

团队讨论

| 导语 相关不等于因果. 本文尝试用通俗的语言介绍因果推断的思想以及因果推断的一种实现方式.

团队文章

- 一. 为什么要使用因果推断
- 二. 因果推断简介

首页

- 三.一种做因果推断的思想

• 参考资料

在过去的研究和分析中,通常使用的机器学习方法的一般做法是将全部的自变量引入模型后对

一. 为什么要使用因果推断

因变量进行预测. 这样的方法可以取得很好的预测效果,但是对于一些预测之外的具体的问题,诸如

• 某个自变量的变化如何影响因变量;

- 两个变量之间是否存在因果关系;
- 若想让因变量朝着自己期待的方向改变,应该主要调整哪个自变量
- 等难以回答和解释.

对于以上三个问题, 比较普遍的做法是:

准确的. 理由如下:

• 若想要研究的自变量X为数值型变量,则绘制出X与因变量Y之间的二维散点图, • 相应地可以计算出二者之间的相关系数,以散点图的走势或相关系数的正负和大小作为二者

- 之间关系的衡量;
- 或对于回归问题直接使用回归系数衡量自变量对因变量的影响大小; • 若想要研究的自变量X为分类型变量,则通过比较E(Y|X=x_i)研究X的不同取值对Y的影响.
- 但上述做法一方面无法回答因果关系的问题,另一方面由该种方法得到的相关关系可能也是不

• 当自变量较多时,对于同一个具体的问题,多重共线性很难避免,换言之即自变量之间的信息 可能会有重叠.当自变量之间存在多重共线性时,由边缘分布直接得出的相关性并不代表真正

- 的相关关系: • 可能会存在一些中间变量使得得到的X与Y之间的相关关系是虚假的; • 相关≠具有因果关系.即使可以得到X与Y之间真正的相关关系,也无法说明X与Y具有因果关 系. 从理论上来讲,**若可以得到**X与Y的关系Y=aX+b,那么**反过来**就可以**得到**X=cY+d,此时难
- 以说明是否具有因果关系或谁是因谁是果.
- 在实际业务中,不仅仅要做出精准预测,还要对预测的结果进行合理的解释,并希望了解某一个变 量的变化如何影响因变量的变化.此时需要对因果关系进行探究,而不应局限于相关关系.

二. 因果推断简介

相关≠因果,如果希望从过去的**非实验数据**中判断某两个指标之间是否存在**因果关系**,直接通过观 察得到结论是并不严谨的. 大家可能非常熟悉这样的挑战——

能会问:

观察发现:收到的互动越多,用户越活跃,因此提出能促进互动增长的某些意见,但其他人可

会不会是本来收到互动多的人就是很活跃(即使没有那么高的互动他们也依然会很活跃),

会被提出这样的问题,本质上是由于收到互动多/收到互动少的两类人群,除了在互动表现上不同, 在其他指标上的表现也不完全相同,这时就没办法确定,到底是互动影响了未来的活跃,还是其他

指标呢? 当然,如果可以寻找到两拨用户,他们除了互动表现不同,其他指标完全同质,自然可以解决这个问 题.但这非常难以实现,因为即使保证了两部分用户的性别,年龄,好友数等等都非常接近,也很难

保证他们在其他没有观测到/没有引入的指标上是完全相同的. 那这样说来,在不能做实验的情况下,因果关系的判断就没救了吗?

一个简单的例子:

或者是更活跃的人才会收到更多的互动(而不是反过来)呢?

并不是!

天上下雨,地面会湿,强相关.那怎么判断是不是因为地面湿了天才下雨呢? 很简单,只要往地上泼水就可以了.如果因为地面湿了才会下雨,那么每一次通过"泼水"这个

升起的原因.

不是下雨的原因. 同样地,公鸡打鸣太阳升起强相关,那么是不是因为公鸡打鸣太阳才会升起呢? 在天黑时让公鸡打鸣,观察太阳会不会升起;或者让所有的公鸡都闭嘴,看看明天太阳会不 会不升起.如果公鸡的声音在或不在,太阳都会照常升起的话,那么公鸡打鸣就并不是太阳

操作把地面弄湿之后,都会下雨.如果泼水或不泼水都不能影响下雨,那么可以认为地面湿

可以看出,一个判断两个事件之间是否存在因果关系的方法就是"干预"! 通过人为地让某个指标产生变化(比如把地面弄湿),观察该指标的变化是否会引起我们关注的y 的变化(观察是否会下雨).如果每次干预都不能影响y,那么可以认为这个指标并不是影响y的原因

如果记某个指标Z对我们关注的Y的平均处理效应为ATE,那么 $ATE:=E(Y \mid Z=1)-E(Y \mid Z=0)$ "干预"这一概念对因果推断十分重要,这里尝试对"干预"进行更加通俗易懂的解释. 更多有关因果

推断的理论知识和术语可参考@elsielin的因果推断文集,珠玉在前,我这里就不再详细介绍啦~

而通过**干预**某指标**导致**的**y的变化**,就可以认为是这个**指标对y的处理效应**.如果怎么干预y都不

变,可以认为处理效应为0,无因果关系;如果干预后y有变化,可以计算变化的大小和方向,从而得

三.一种做因果推断的思想

已经不可能进行了,也就是说,一个Z=1的人,没办法把他过去的Z置为0再观察y的结果.一个已经 上了大学的人,没办法让他回到过去选择不上大学,来观察后期的职业发展是否会有什么不同.

之一,反之则可以得到具有因果关系的结论.

出相应的因果关系推断的结论.

那怎么办呢? 没有办法回到过去真实地干预某个人,但我们可以对他的"数据"进行"干预".

换句话说,如果我们可以训练一个模型,能够同时预测出一个人的Z=1和Z=0时y的表现,将两个结

uin

uin

...

果相减,就是处理变量Z对这一个人的处理效应,全部人做平均就是处理变量的平均处理效应.

"干预"的方法虽好,可如果是用历史数据进行因果推断,所有事情都已经发生完毕,"干预"某个指标

更加具体地, • 通过已有的真实数据训练一个预测模型;

• 再将全部人的Z都置为0,用模型对y进行预测; • 将两次预测的结果相减求平均,就是变量Z对y的平均处理效应.

114

92

80

真实的Y uin

0

如果平均处理效应显著,则认为有因果关系.

• 之后将全部人的Z都置为1,用模型对y进行预测;

	4		0	98	
	5		1	110	
	6		1	82	
				[因	果推断
7	传统的回归	模型都可以	人实现上述	步骤,比如	随机森林

	4		1	101.2			
	5		1	110.1			
	6		1	80.8			
[因果推断流	程]						
传统的回归模型都可以实现上述步骤,比如随机森林等等,下篇文章主要介绍一种需要借助贝叶							
斯先验和MCMC的贝叶斯累加回归树BART方法.							
参考资料							

预测的Y0

116.6

90.5

96.2

107.1

74.8

预测的Y1

124.6

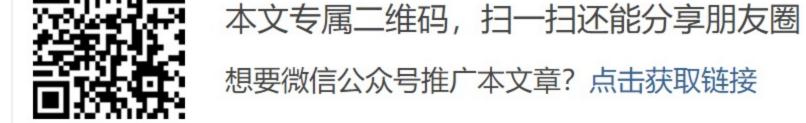
94.5

of-Statistics-and-Machine-Learning Jennifer-Hill.pdf 最后更新于 2020-07-29 18:48

仅供内部学习与交流,未经公司授权切勿外传

2. https://nyhackr.blob.core.windows.net/presentations/Causal-Inference-at-the-Intersection-

1. https://iyarlin.github.io/



标签: 因果推断;

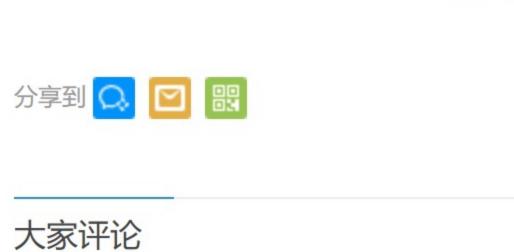
想要微信公众号推广本文章? 点击获取链接

• 因果推断简介(二): 倾向性得分匹配介绍

相关阅读

- 图果推断实战(一):微信看一看"猜涨跌"小结
- 因果推断学习笔记: ACIC Data Challenge 学习笔记 • 因果推断简介(三): 倾向性得分加权介绍

我顶 (6)

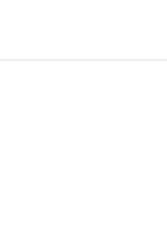


elsielin

期待~

▲顶 早回复

• 因果推断简介(一): 掀开"因果推断"的面纱



收藏 (27)

己 转载

Copyright©1998-2021 Tencent Inc. All Rights Reserved

腾讯公司研发管理部 版权所有

广告申请 反馈问题

[1951/1979/332 ms]

Ø 收录 ○ 评论 (1) □ 反馈

选中文章内容可快速反馈

2019-08-27 14:33:59

切换到更多功能 发表评论

关于作者

团队投票

WXG\基础产品部\数据分析一组员工

xuenanwang(王雪楠) 🔀

团队日历

团队会议

作者文章 Bye~

一种基于selection bias的因果推断敏感性分析...

- 因果推断简介(四): 因果推断实战

• 贝叶斯结构时间序列因果推断方法

因果推断简介(三): 敏感性分析

离线因果推断简介

浏览 3929 收藏 226 因果推断读书笔记

收录于

猜你喜欢

【晋级PPT】游戏策划晋升P10-Christianlu(陆...

更多>>

- 【晋级PPT】游戏策划晋升P10-Chandleyang 【晋级PPT】游戏客户端开发晋升T10-Lvyhe(...
- 【晋级PPT】后台开发晋升T11-Pizhou(周原)