

Donald Rubin亲自讲解什么是因果推断 | 因果科学读书会第1期

【读书会相关信息】

“因果”并不是一个新概念，而是一个已经在多个学科中使用了数十年的分析技术。通过前两季分享，我们主要梳理了因果科学在计算机领域的前沿进展。如要融会贯通，我们需要回顾数十年来在社会学、经济学、医学、生物学等多个领域中，都是使用了什么样的因果模型、以什么样的范式、解决了什么样的问题。我们还要尝试进行对比和创新，看能否以现在的眼光，用其他的模型，为这些研究提供新的解决思路。

如果大家对这个读书会感兴趣，欢迎报名：

https://pattern.swarma.org/mobile/study_group/10?from=wechat

【时间】2021年10月22日 21:00--23:00

【主讲人】

Donald B. Rubin教授，美国国家科学院院士，美国科学与艺术学院院士，曾任哈佛大学统计系John L. Loeb讲席教授，现任清华大学丘成桐数学科学中心教授。他获得过统计学领域几乎所有著名奖项，包括著名的 Woilks奖章、Parzen奖、Snedecor奖等，是当今世界最具影响力的统计学泰斗。他对科学的贡献已超出统计学范畴，其统计思想对生物医学、经济学、心理学、教育学、社会学及计算机科学等众多领域产生了划时代的影响，谷歌学术显示其文章和著作引用量已超过35万次。

【笔记小分队】

邓宇昊 北京大学数学科学学院博士生

刘 青

【讲座笔记】

【记录人】刘青

【主讲人】Donald B. Rubin

本期内容

随机实验和观察研究中因果推理的基本概念：一段非凡的历史。

1. 解读了潜在结果模型以及基本概念，包括行为、单位、干预、分配机制、观察结果、潜在结果、随机对照实验。
2. 潜在结果指接受潜在治疗的个体的结果，通常在有限样本中，视每个个体的潜在结果是命定的，而随机性来源于分配机制（Assignment Mechanisms）。若分配机制已知，则称为实验性研究（Experimental Studies），否则称为观察性研究（Observational Studies）。

3. 意识是人类的特长。
4. 干预：为了了解科学，你所做的就是以某种方式介入现实世界，以测量世界科学的各个方面，并在某个时间点进行干预。
5. 海森堡不确定性原理：你不可能同时知道一个粒子的位置和它的速度，粒子位置的不确定性，必然大于或等于普朗克常数（Planck constant）除以 4π （ $\Delta x \Delta p \geq h/4\pi$ ），这表明微观世界的粒子行为与宏观物质很不一样。
6. 观察者效应，即积极的观察实际上改变了世界，与当时的科学状态有关，与之前的科学状态不同。海森堡测不准原理在哪里。它涉及到量子力学以及在同一时间点测量两个物体的不确定性。
7. 因果推断最根本的问题是丢失数据问题。
8. 随机化：通过随机化了解一种干预措施实施的结果。
9. 估计值：随机实验中要估计的数量，作为单位内潜在结果的函数。
10. 置信区间：置信区间是指由样本统计量所构造的总体参数的估计区间。
11. 随机试验：是在相同条件下对某随机现象进行的大量重复观测。开展统计分析的基础。概率统计需要对某随机现象进行大量的重复观测，或在相同条件下重复试验，观察其结果，才能获得统计规律性的认识。非随机试验。
12. 潜在结果，由观察结果代替，每个治疗单元的指标 w_i ，因此他们将对观察值进行回归，这就是为什么 i 可能性，作为指标变量 w_i 和协方差上每个单元的观察。
13. 必须模拟分配机制来推断缺失的潜在结果。通过分配机制模型，创建缺失的和观察到的数据，以便对缺失的潜在结果进行推断----这是今年诺贝尔奖背后的基本思想。
14. 给出了目前潜在结果模型的一些结论。保持过去关键的见解；避免过去的混乱；意识到RCT的想法是非常新的；更新设计和分析方法，以利用现代计算机环境；鼓励数学上的精确，尤其是在逻辑流程上；精确度可能会对具有挑战性的应用产生严重后果。
15. 一些更新的一般的想法。

【精彩问答】

周晓华：首先，我对Rubin因果模型做一些评述。Rubin因果模型对待估量和模型（估计方法）做出了清晰的界定，这一点在研究中非常重要。待估量表示的是科学问题，在有明确的待估量之后，再考虑用什么模型、用什么方法估计，以及要做出什么样的假设才能识别出参数。现在有很多人人工智能的文章，混淆了待估量和模型，我相信Rubin的报告会给相关领域的研究提供很大启发。

周晓华：我想补充一些关于主分层（principal stratification）的应用。在一些临床试验中，一些个体会死亡，如果我们关心的是患者接受治疗一年后的生活质量，而患者在接受治疗一年内死亡了，那么他的生活质量是无法定义的，这被称作“死亡截断”（truncation by death）。一些人把死亡截断当作缺失处理，但这是不对的。正确的做法是用主分层。

周晓华：我还想提一下精准医疗（precision medicine）的问题。也许一种药无法对所有人都有效，所以精准医疗也被叫做个性化治疗。我们关心的不是整个人群上的平均因果作用，而是子人

群（也就是协变量取某个特定值）的因果作用。精准医疗的核心是如何处理异质性的因果作用，特别是如何用观察性数据做决策，因为观察性数据中包含未观测到的混杂。现在因果研究的一大挑战是，如何把Rubin因果模型应用于更复杂的情形。

Rubin：的确，死亡截断是一个重要的问题。如果想研究某种干预措施对患者生活质量的影响，而患者死亡了，我们需要做一些工作，把死亡和感兴趣的结局分开。如果简单地把死亡的个体扔掉，我们会得到完全错误的结论，比如我们可能会发现治疗会杀死弱者，但这不能代表治疗对生活质量的的作用。所以，要用主分层定义一个永远幸存者的群体，在这个群体上考虑因果作用。关于精准医疗，我的观点是，如果对不同人有不同的治疗，主层只能部分识别，这时协变量就至关重要了，协变量能够帮助预测一个个体处于哪个主层。

崔鹏：近些年来，机器学习领域的学者越来越关心因果的问题了。一些学者开始反思机器学习的局限性，比如缺乏稳定性、缺乏泛化性。当我们在预测问题上引入因果分析，我们能做些什么？

Rubin：我们首先要确定，要回答的问题是什么，干预是什么。二十世纪的一个重要统计思想是试验设计，然而在机器学习领域，我们能施加哪些干预？计算机网络中的干预要比经典的干预复杂得多，因为个体之间会互相影响。即使能做随机化试验，也很难找到独立的个体单元。所以我的建议是，首先明确能施加什么干预，其次如何处理个体间的交互。特别是，实际问题可能会遇到网络结构，如果处理网络上的因果分析还有大量工作要做。

崔鹏：我的理解是，机器学习的核心任务是预测。在经典的设定中，我们假设个体是独立同分布的，另一个设定是域适应（domain adaption），有一个训练集，还有另一个任务集，任务集的特征已知。机器学习要做的是拟合任务集上的协变量分布。但更复杂的设定是，我们不知道任务集的特征，如何调整协变量的分布，才能做出正确的预测呢？如果我们有数据生成过程的正确模型，那么显然我们能在各种任务集上都达到一致好的预测表现。反过来，当改变数据的分布，如果一个模型能达到一致好的预测表现，就可以认为这个模型很好地捕捉了数据的特征。所以我认为，在机器学习中，干预也可以通过对协变量再加权来实现。

Rubin：如果协变量很简单，再加权确实起作用。但如果协变量很复杂，而你又有很多结局，这时你就需要模型了，用模型来刻画结局如何依赖于协变量。如果你有一个好的模型，那么你就能用协变量预测结局了。为了更好地建模，就需要加权、甚至是连续版本的加权。我相信我们说的同一件事，尽管语言上有所差异。在二十世纪三十年代的抽样调查研究中，也许和你说的计算机科学有些相似，收集到的数据有缺失，缺失个体的特征可能有很大不同，为了外推就需要假想的模型，但加权可以在缺失个体和不缺失个体的相同支撑集上才能做出预测。有文献总结过如何处理缺失数据，其中可能有些有趣的想法。

赵西亮：计量经济学经常使用自然试验来识别因果作用，Rubin因果模型非常适合回答政策评估的问题，通过干预，就能看出政策的因果作用。计量经济学的另一问题是，我们观察到某个现象，而我们想知道发生这一现象的原因是什么。Rubin因果模型如何解决这一问题呢？

Rubin：这是一个非常重要的问题。当你查阅大量因果推断的文献，就会发现有很多人都试图寻找一个结果的原因。在随机化试验中，通过随机化“原因”，就能观察到各个“原因”的结果。但对于观察性的数据，情况就变得复杂了。不妨举一个简单的例子。假设现在某人死于肺癌，一个人说，他死于肺癌是因为他吸烟，另一个人说，他吸烟的原因是他父母吸烟，第三个人说，他父母吸烟又是因为他的祖父母吸烟，这样可以无穷无尽追溯上去。我认为，二十世纪的一大革命是试验设计，从寻找原因转向到干预（分配机制）。在实际研究中，我们要关注我们能够理解的

问题，注意我们能做什么干预，然后考察这些现实的干预的结果，诸如“把太阳从太阳系中移除会发生什么”、“如果我能活到两千岁”这种问题是没有意义的。

观众：你对潜在结果有何期待？

Rubin：潜在结果模型是正确的而且简单的框架，但在实际问题中还有很大的完善空间，比如周晓华提到的精准医疗。当存在协变量的时候，如果要做随机化试验，直接忽视协变量吗？如果你只是做简单的试验，这未尝不可，你不必在意协变量的不平衡，但如果你想做决策，忽视协变量就不可能达到“精准”，因为决策要基于“条件”的量。即使我们具有高效的计算工具来进行试验设计，一个重要的观点是再随机化（rerandomization）。对于再随机化，你要考察处理组和对照组的协变量分布，把协变量不平衡的分配方案抛弃掉。把不合理的分配抛弃掉，协变量分布的差异就不再服从常规的t分布（或卡方分布）了，而是一些截断分布的组合。用潜在结果的一般框架能够改善因果模型。Fisher曾说过，他会把不合理的随机化舍弃掉。比如，长势好的作物被分配到了某种处理，长势不好的作物被分配到了另一种处理，这样的分配就不合理。事实上，Fisher发明了协方差分析就是为了校正这一问题。

观众：如何用因果的方法解决解决流行病学的问题？

Rubin：以疫苗为例，美国的政策是以州为单位发布的，这为我们比较不同的政策提供了方便。现实世界的各种政策提供了大量的“自然试验”，其中可能存在某些分配机制。分配机制也许和潜在结果无关，这是一个很关键的假设，在分析中需要认识到这一点。

观众：潜在结果和分配机制有什么关系？如何设计更好的试验？

Rubin：潜在结果描述的是科学问题，它和分配机制是无关的。我在二十世纪七十年代做出的一项重要贡献就是强调潜在结果框架和干预的区别，前者定义了因果量，后者是关于如何测量这一因果量。分配机制描述的是你如何回答科学问题，分配机制可以是来自随机化的、来自于自然试验的，或者来自于观察行研究。在解决实际问题时，要先要把你关心的问题用潜在结果的语言表述出来，然后思考你能够对个体施加怎样的干预。对于复杂的试验设计，比如嵌套或集群随机化试验，仔细想想，你观察到的是什么潜在结果，缺失的是什么潜在结果？我比较喜欢建立贝叶斯模型，从随机化或假象的分布开始，然后利用观察到的数据，得出更细致的结论。

【参考文献】



Essential concepts of causal inference.pdf

1.48MB



【读书会PPT】



Essential.Concepts.for.Causal.Inference-a remarkable history-_25_Sept_.2020.pdf

85.02KB



