

## 因果推断: 起源和发展

王东明, 陈都鑫

(东南大学 数学学院, 江苏 南京 210000)



**摘要:**近年来,随着因果理论的逐渐完善和大数据时代的到来,因果推断已经发展为一个跨学科的重要研究课题。主要综述了因果理论的一些研究进展,首先介绍了因果理论有重要影响的结构化因果框架和潜在结果框架,分析了二者的异同,并通过实例演示了一般因果推断的过程;然后分别介绍了在不同学科领域衍生的线性、非线性和部分线性因果模型的重要理论成果;接下来介绍了基于数据的因果推断和因果网络构建算法;最后总结了因果推断可能的发展方向和对机器学习的影响。

**关键词:**因果推断;因果模型;因果网络;人工智能

**中图分类号:** TP3

**文献标识码:** A

## Causal Inference: Origin and Development

WANG Dong-ming, CHEN Du-xin

(School of Mathematics, Southeast University, Nanjing, China, 210096)

**Abstract:** In recent years, with the gradual improvement of causal theory and the advent of the era of big data, causal inference has developed into an important research topic across many disciplines. The research progress of causality theory is summarized in this paper. Firstly, it focuses on the structured causality framework and potential result framework that have an important influence on causality theory. The similarities and differences between the two are analyzed, and the process of general causal inference is demonstrated by examples. Next, it introduces important theoretical results of the linearity, non-linearity, and partial-linearity frameworks derived from different disciplines. Subsequently, data-based causal inference and causal network construction algorithms are introduced. Finally, the possible development direction of causal inference and its impact on machine learning are summarized.

**Key words:** Causal inference; causal model; causal network; artificial intelligence

### 1 引言

自相关系数的概念被提出以来,“涉及因果推断的问题自始就缠住了统计学的脚后跟”<sup>[1]</sup>。在因果的概念被明确界定并广泛使用之前,相关性和因果性常常是互换使用的,但是不可忽略的是二者有着本质的不同。关于因果是什么,曾经在很长的一段时间内都是一个经验的问题,在文献[2,3]中,Granger 这样定义因果:① 原因应该发生在结果之前;② 原因应包含其他方式无法获得的原因信息。可以结合相关性的概念这样理解这一定义:通常而言,相关性表示一般的关联关系,表现为两个变量的变化在统计意义上是互相联系的<sup>[4]</sup>。因果关联则更加倾向于结果依赖或部分依赖于原因,而原因造

成或共同造成结果的发生。相较于相关性而言,因果性的一个重要的特点即为关系的方向性,关联性是双向的,而因果性通常是单向的。根据因果性的特点,因果推理即为根据某一效应发生的条件,对某一因果关系作结论的过程。因果推理与关联推理的主要区别在于,前者分析的是当原因发生变化时,效果变量的响应,而后者仅仅强调二者之间是否相关<sup>[5,6]</sup>。

本文主要对因果推断和因果理论的发展进行了总结。第2节主要介绍了因果推断现有的理论模型和在因果推断领域影响最为广泛的结构化因果模型和潜在结果模型;第3节主要介绍了从模型的角度

收稿日期: 2021-06-30; 修回日期: 2021-07-13

基金项目: 国家自然科学基金资助项目(61903079)

作者简介: 王东明(1997-),男,河南济源人,研究生,主要研究方向为因果推断和因果模型等;陈都鑫(1989-),男,江苏徐州人,博士,讲师,主要从事群体智能信息挖掘与分布式控制等方面的教学与科研工作(本文通信作者, Email: chendx@seu.edu.cn)。

度出发,考虑因果演化过程的基于模型的因果推断方法,主要分类介绍了线性模型、非线性模型和半线性模型的因果推断方法;第4节介绍了在数据结构分析中发挥了重要作用的基于时间序列数据的因果推断方法,以及该领域目前理论较为完善的最优因果熵原理和两步法因果推断方法。值得一提的是,许多线性或非线性模型在应用于高维时间序列因果推断时也有非常不错的效果,但是由于其考虑问题的角度和场景不同,在本文中也将对其分别综述;第5节总结了因果推断的发展现状和仍未解决的问题,主要分析了因果推断在机器学习和人工智能领域仍未被发掘的潜力和可能的发展方向。

## 2 两类经典因果推断模型

在因果推断快速发展的几十年里,不同学科背景下产生了众多的推断方法,如基于匹配的因果推断方法<sup>[7]</sup>,基于树和因果集合的因果推断方法<sup>[8]</sup>,以及动态扰动下的因果推断方法<sup>[9]</sup>。但这些方法大多应用局限而没有给出一个一般性的理论结构,因此产生的影响比较有限。

在实际问题中,可以根据一般经验判断和确定一些简单的因果关系,然而当变量众多、问题复杂时,判断因果关联就会变成一个非常困难的事情。推断因果关系最有效的方法是进行随机对照实验,将所有实验对象随机分配到干预组和对照组。设计实验过程,使对照组和治疗组之间的唯一预期差异是正在研究的结果变量。然而,在现实中,随机对照实验受限于实验成本而无法实现。此外,随机对照实验只关注样本的平均情况,并没有解释个体机制或区别单个实验样本。大多数随机对照实验还需要考虑伦理问题,这在很大程度上限制了其应用。根据原始数据构造因果推断方法而不依赖于随机实验是一个非常重要的研究课题。观测数据是研究者在不进行任何干扰的情况下对被试者进行简单地观察而获得的。因此,研究人员无法控制样本的潜在变量,可以理解为仅仅根据观察记录数据,从观测数据中,可以找到观测目标的行动、结果,以及事件演化过程。但他们为什么采取特定行动的机制仍然是未知的。对于观测数据,核心问题是如何得到与事实相反的结果。例如,如果病人接受不同的药物治疗是否会有不同的结果?一般而言,回答这些反事实的问题并不容易。其一,因为只观察到事实的结果,而从来没有观察到如果他们选择了不同的治疗方案可能会发生的反事实的结果。其二,因为在观察数据中,“治疗”这一事件在人群中常常不是随机的,诸多不可观测的潜在变量,例如经济水平、

患者个人身体情况等,可能导致受治疗人群与一般人群产生差异<sup>[10]</sup>。

为了解决观察数据的因果推理中的这些问题,目前已经有许多因果模型被提出。其中理论相对完善、应用较为广泛、在因果推断领域认可度比较高的有潜在结果模型<sup>[11,12]</sup>和结构化因果模型<sup>[13~15]</sup>。潜在结果模型也被称为Neyman-Rubin潜在结果模型或Rubin因果模型。结构化因果模型(SCM)框架包括因果图、结构化方程、反事实和介入式逻辑。区别于潜在因果模型,结构化因果模型侧重于描述一个系统的因果机制,通过一组结构方程来构造随机变量间的因果关联。

本节主要介绍潜在结果模型框架和结构化因果模型框架<sup>[5,16]</sup>的理论背景、统计和机器学习领域的代表性方法,最终介绍了这两种不同框架的区别和联系。

### 2.1 潜在结果模型框架

潜在结果模型的核心思想可以用一个简单的例子来说明:一个学生每天正常地吃早餐,对他的体重进行观测;而如果没有正常地吃早餐,自然地认为他应当有一个不同的体重。为了衡量吃早餐与体重变化之间的因果效应,需要比较同一个人在两种情况下的结果。显然,同时观测到两种潜在的结果是不可能的,其中一种潜在的结果总是缺失的。潜在结果模型旨在评估这些潜在结果,然后得到二者之间的因果相关关系。因此,在潜在结果框架下,对因果效应的估计是因果推理的核心问题之一。

关于潜在结果模型的起源,目前仍未有一个确定性的定论。通常认为,Rubin<sup>[17]</sup>在潜在结果理论模型的发展过程中做出了革命性贡献,这可能也是潜在结果模型被命名为Rubin因果模型的原因之一。但在Rubin提出潜在结果模型之前,在不同领域有众多学者已经对这一概念进行过一些探索。在对Rubin提出这一理论的过程进行调研后发现,统计学和哲学是潜在结果模型的重要思想来源。其中Hume<sup>[18]</sup>和Mill<sup>[19]</sup>是哲学家的典型代表,他们最早讨论反事实框架,并给出相关定义;Neyman<sup>[20]</sup>和Fisher<sup>[21,22]</sup>从统计学的角度分别提出从潜在结果模型的角度来讨论随机变量间的因果关系。Heckman<sup>[23]</sup>认为,Roy也曾曾在经济学领域提到过潜在结果模型的思想。综上所述,首次提出完整潜在结果模型框架的仍然是Rubin,所以他是潜在结果模型的主要发现者和开创者。

潜在结果模型的核心是比较同一个研究对象在接受干预和不接受干预时的结果差异,认为这一

结果差异就是接受干预相对于不接受干预的效果。对同一研究对象而言,通常不能够既观察其干预的结果,又观察其不干预的结果。对于接受干预的研究对象而言,不接受干预时的状态是一种“反事实”状态;对于不接受干预的研究对象而言,接受干预时的状态也是一种“反事实”状态;所以该模型又被某些研究者称为反事实框架。但是,Rubin并不认同“反事实框架”的概念,他认为结果的出现与否主要取决于干预机制,这并不意味着一种结果不存在,只是事实上只能看到一种结果。Rubin认为潜在结果是一个更合适的概念。考虑到这一理论主要是由Rubin开创并持续推动的,在介绍该理论时,也使用了潜在结果模型的概念。下面,首先给出潜在因果模型的一些理论。

首先给出潜在结果框架下的符号定义<sup>[11,24]</sup>,它在逻辑上等价于结构化因果模型框架<sup>[25]</sup>。潜在结果框架的基础是因果关系与干预相关联<sup>[26]</sup>。通过比较研究对象在不同干预下的结果差异得到干预变量和结果之间的因果关联。下面,首先介绍因果推理中的三个基本概念:单元、干预和结果。

#### 定义1 单元

单元是干预效果研究中的原子(指最小)研究对象。

一个单元可以是一个物理对象,或者是一个物理对象或人的集合,如教室或市场<sup>[26]</sup>。在潜在结果框架下,不同时间点的单元是不同的。数据集中的一个单元是整个随机变量对应的样本,因此在潜在结果模型下,术语“样本”和“单元”可互换使用。

#### 定义2 干预

干预即为对一个单元采取的扰动。

记 $W(W \in \{0,1,\dots,N_w\})$ 表示干预,其中 $N_w+1$ 是可能干预的总数。例如在前面提到的例子中,吃早餐与不吃早餐即为一个简单的二元干预。现有的干预大多数都为二元干预,记干预变量 $W=1$ 的单元为干预组,干预 $W=0$ 的单元组为对照组。

#### 定义3 潜在结果

对于每个单元-干预对,当干预应用于单元时,产生的结果是潜在结果<sup>[26]</sup>。

干预 $w$ 的潜在结果表示为 $Y(W=w)$ 。

#### 定义4 观测结果

观测结果是实际情形下的结果。

观测结果也称为事实结果,用 $Y^F$ 来表示。潜在结果和观测结果之间的关系为 $Y^F = Y(W=w)$ ,其中, $w$ 对应于实际情形下的干预变量取值。

#### 定义5 反事实结果

反事实结果是如果该单元采取了另一种干预时

产生的结果。

反事实结果是干预的潜在结果。由于一个单元只能接受一种干预,所以只能观察到一种潜在的结果,其余未观察到的潜在结果是反事实结果。在多重干预的情况下,用 $Y^{CF}(W=w')$ 表示干预的反事实结果。在二元干预的情况下,可以直接记:

$$Y^{CF} = Y(W=1-w) \quad (1)$$

式中, $w$ 即为实际采取的干预。

在观察数据中,除了选择的干预和观察的结果外,还记录了单元在其他干预下的信息,这些信息可以用于区分干预前变量和干预后变量。

#### 定义6 干预前变量

干预前变量是不会受到干预影响的变量。

干预前变量也称为背景变量,它们可以是患者的数量、病史等。记 $X$ 为干预前变量。

#### 定义7 干预后变量

干预后变量是受干预影响的变量。

在以下章节中,变量是指干预前变量,除非另有说明。

在介绍了上述概念后,可以对干预效果进行定量定义。首先给出二元干预下的干预效果定义,并且可以通过比较它们的潜在结果将其扩展到多元干预的情形。

在总体层面上,干预效果定义为整个群体中的平均干预效果(ATE),如下:

$$ATE = E(Y(W=1) - Y(W=0)) \quad (2)$$

式中, $Y(W=1)$ 和 $Y(W=0)$ 分别是整个群体的潜在干预和非干预结果。

对于干预组,干预效果定义为干预组的平均干预效果(ATT),如下:

$$ATT = E(Y(W=1)|W=1) - E(Y(W=0)|W=0) \quad (3)$$

式中, $Y(W=1)|W=1$ 和 $Y(W=0)|W=0$ 分别是干预组的潜在干预和对照结果。

在子集层面上,干预效果定义为条件平均干预效果(CATE),如下:

$$CATE = E[Y(W=1)|X=x] - E[Y(W=0)|X=x] \quad (4)$$

式中, $Y(W=1)|X=x$ 和 $Y(W=0)|X=x$ 分别是 $X=x$ 的子集的潜在干预和对照结果。CATE是在干预效果在不同子集之间变化的情况下的常见干预效果,也称为异质性干预效果。

在个体水平上,干预效果被称为个体干预效果(ITE),单元 $i$ 的ITE被定义为

$$ITE_i = Y_i(W=1) - Y_i(W=0) \quad (5)$$

式中,  $Y_i(W=1)$  和  $Y_i(W=0)$  分别是单元  $i$  的潜在干预和控制结果。在一些文献<sup>[27,28]</sup>中, ITE 被视为等同于 CATE。

对于因果推断, 目标是从观察数据中估计干预效果。从形式上看, 给定观测数据集  $\{X_i, W_i, Y_i^F\}_{i=1}^N$ , 其中,  $N$  是数据集中的单元总数, 潜在因果模型框架下因果推断任务的目标是估计上述定义的干预效果。

## 2.2 结构化因果模型框架

结构化因果模型(SCM)最早由 Judea 在 1995 年提出并在 2000 年进一步完善<sup>[13,29]</sup>, 该模型结合了经济学和社会科学中使用的结构方程模型(SEM)的特征<sup>[30]</sup>, Neyman(1923)和 Rubin(1974)的潜在结果框架, 以及概率推理和因果分析理论框架下的图形模型<sup>[31-33]</sup>。Judea 将结构化因果模型的产生盛赞为“因果革命”。结构化因果革命提供了一种从图结构进行因果发掘的思路, 通过包括 d-分离准则、前门准则、后门准则等基本理论和概率图模型相关的基础理论构建出了一套基于反事实推断的因果推断方法。

抛开结构化因果模型复杂的理论背景, 可以认为结构化因果模型由三部分组成: ① 图模型; ② 结构方程; ③ 反事实干预。三者在结构化因果模型中起到的作用为图形模型作为一种语言表示先验经验数据结构, 反事实用于阐明问题, 结构方程则倾向于利用数学的语言将二者结合起来。

以推理机的形式给出了结构化因果模型的模型框架, 如图 1 所示。

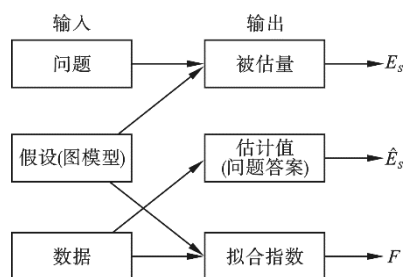


图1 结构化因果模型数据-模型结合机理（推理机）  
Fig. 1 Data-model combination mechanism of structured causal model (inference engine)

总体来看, 结构化因果模型根据三个输入, 假设、问题和数据。生成三个输出, 被估量、估计值和拟合指数。被估量是一个基于假设的数学公式, 它给出一个从数据中生成问题答案的策略。接收到数据后, 使用该估计策略生成答案的实际估计, 以及对被估量置信度的统计估计(反映数据的缺失等其他因素)。最后, 生成一个“拟合指数”列表, 该

列表衡量数据与模型所传达假设的兼容性。

下面通过一个例子来说明结构化因果模型的机理, 假设问题为估计随机变量  $X$  对随机变量  $Y$  的因果影响, 记作  $Q = P(Y|\text{do}(X))$ 。假设其因果图如图 2 所示。

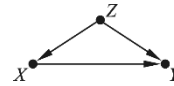


图2 关于随机变量  $X, Y, Z$  的因果图  
Fig. 2 Causal graph of random variables  $X, Y, Z$

图 2 中,  $Z$  是影响  $X$  和  $Y$  的第 3 个随机变量。最后, 在服从联合分布  $P(X, Y, Z)$  的数据中随机采样。估算值  $E_s$  为

$$E_s = \sum_Z P(Y|X, Z)P(Z) \quad (6)$$

该式的具体推导本文不再赘述。根据结构化因果模型机理, 可以知道式(6)为随机变量  $X$  对随机变量  $Y$  的因果影响  $Q$  提供了一个解答方案。而根据样本数据对式(6)进行估计即可得到一个对应的估计值, 这里统计学和图模型理论中有很多方法可以做到一致性估计。例如, 满足指定的  $X$  和  $Z$  条件的所有情况下的样本平均值就是一个一致的估计, 这一点可以由格里汶科定理保证。但是当数据量较少、数据较为稀疏时, 传统统计学的方法未必准确, 这时就需要设计更有效的估计方法来克服数据稀疏性<sup>[34]</sup>。这是深度学习的优势所在, 也是大多数机器学习工作的重点所在。最终, 在做出判断后, 拟合指标为空。换句话说, 在检查了图的结构之后, 编码的假设就没有任何可测试的含义。因此, 最终估计的准确性必须完全依赖于图表中编码的假设。

同样的过程适用于更复杂的问题。例如, 可以考虑更多变量对同一变量是否存在因果关联, 进一步可以考虑两个变量集之间的因果关系。发现这样的因果结构也可以允许一些数据从受控实验中得到, 可以采取  $P(V|\text{do}(W))$  的形式生成受控变量  $W$  的数据, 估计要求仍然是将问题转化为已有数据可以回答的格式, 然后采用一些估计手段进行估计, 以确保估计的无偏性。但是受限于数据等诸多因素, 问题并不总是可回答的, 针对这一问题 Bareinboim 和 Pearl 开发了一套高效的算法来判断可识别性, 并针对各种反事实问题和各种数据类型提供可估计行条件和可识别性要求<sup>[35]</sup>。

值得一提的是, 部分变量不可观测, 当这些变量是影响结果的重要变量时, 通常想要通过数据准

确估计因果效应就变得困难甚至不可能实现,此时有一些工作针对部分变量不可观测的情况给出了因果关系的一个区间估计的方法,其中 Li 和 Pearl 在这一方面做出了重要贡献。至此,结构化因果模型的理论框架已经相对完善。这里给出一个例子来简单阐述结构化因果模型在面对有不可观测的潜在变量时的界估计方法<sup>[36]</sup>。

研究一种服药和康复之间的因果关联。实验包含 700 名患者的恢复率。在所有的患者中,共有 192 名患者选择服用该药物,508 名患者没有服用。血型(是否 O 型)和年龄(是否 70 岁以下)是服用药物和康复的混淆因素。与年龄相关的数据缺失,仅仅知道当地 85.43% 的人年龄在 70 岁以下。因为年龄和血型都是服用药物和康复的混杂因素,并且与年龄相关的观察数据不可观测,所以因果效应是不可识别的。

为了更好地描述数据,我们首先给出因果图,如图 3 所示。考虑血型混杂因素的实验结果(康复率,表中 A/B 表示 B 人中有 A 个康复)和考虑血型-年龄双混杂因素的实验结果分别见表 1 和表 2。

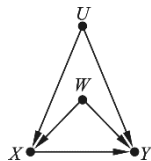


图 3 服药-康复因果结构图

Fig. 3 Taking medicine-rehabilitation causal structure graph

图 3 中,  $X$  表示患者服药情况,取值为 1 表示服药,取值为 0 表示未服药。 $Y$  表示患者康复情况,取值为 1 表示康复,取值为 0 表示未康复。 $U$  和  $W$  是混杂因素, $U$  表示患者年龄是否是 70 岁以下(取值为 1 表示是,取值为 0 表示不是), $W$  表示患者是否是 O 型血(取值为 1 表示是,取值为 0 表示不是)。

表 1 考虑血型混杂因素的实验结果

Tab. 1 Experimental results considering confounding factors of blood type

	$X=1$	$X=0$
$W=1$	23/36	145/225
$W=0$	135/156	152/283
总计	158/192	297/508

表 2 考虑血型和年龄双混杂因素的实验结果

Tab. 2 Experimental results considering both confounding factors of blood type and age

	$X=1$	$X=0$
$W=1, U=1$	3/4	141/219

表 2 续

	$X=1$	$X=0$
$W=1, U=0$	20/32	4/6
$W=0, U=1$	135/151	117/224
$W=0, U=0$	0/5	35/59
总计	158/192	297/508

根据文献[36]中给出的 Tian-Pearl 界我们可以得到对因果效应的估计如下:

$$P(X=1, Y=1) = \sum_w P(Y=1 | X=1, W) \quad (7)$$

$$P(W=1|W)P(W)=0.2257$$

$$1 - P(X=1, Y=0) = 1 - \sum_w P(Y=0 | X=1, W) \quad (8)$$

$$P(X=1|W)P(W)=0.9514$$

由此得到因果效应的界限是  $0.2257 \leq P(Y=1 | \text{do}(X=1)) \leq 0.9514$ , 其中不使用协变量  $W$  的因果信息和先验信息  $P(U)$ 。界限之间的差距(即  $0.9514 - 0.2257 = 0.7257$ )过大,因此这样的估计通常认为是无效的。

在因果图中我们发现  $U \cup W$  满足后门准则,因此可以利用  $P(X, Y, W)$  和  $P(U)$  的信息。借助文献[37]中的“SLSQP”求解器,可以得到因果效应的界,为  $0.4728 \leq P(y | \text{do}(x)) \leq 0.9514$ 。下限显著增加,中点是 0.7121。相对上一种结果而言,这是更能被接受的结果,最终我们得到服药后恢复的因果效应为 0.7121。该数值即表示二者间因果联系的强弱,换言之,结论可以被表述为服药对康复有显著因果关系。

迄今为止,结构化因果模型已经在众多行业得到了广泛的应用,例如在流行病分析领域<sup>[38,39]</sup>、统计领域<sup>[40,41]</sup>和社会科学领域<sup>[42]</sup>。其作为因果关系综合理论的潜力尚未得到充分利用。迄今为止,结构化因果模型的价值主要体现在如下方面。

① 分别应用结构化因果框架各个组成部分:因果图、潜在结果、结构方程、决策分析<sup>[43]</sup>、干预<sup>[44]</sup>、充分性论证<sup>[45]</sup>和概率<sup>[40]</sup>;上述每种方法都可以看作是部分或者受相对应限制的结构化因果框架。

② 反事实和反事实联合概率的定义、公理化和算法化。

③ 将对“原因的影响”、“间接影响”和“结果的原因”的评估转换为算法分析的方法。

④ 从数学(统计学、图论、概率图模型等)的角度解释和完善了潜在结果模型,制定结构方程模型的反事实基础。

⑤ 从数学的角度严谨给出了诸如“反事实”、

“干预”、“混淆”等定义，严格区分了相关性和因果性。

⑥ 修正传统因果观点，给出了更为一般性的数学-逻辑框架<sup>[46-50]</sup>。

### 2.3 小结

本部分主要介绍了结构化因果模型和潜在因果模型的模型架构和模型特点。Pearl 在提出结构化因果模型时借鉴了 Rubin 反事实框架的思想（虽然 Rubin 本人并不认可反事实逻辑，他认为结果的出现与否主要取决于干预机制，这并不意味着一种结果不存在，只是我们只能看到一种结果），而利用图结构信息体现变量间的结构化因果关联并发展了前门、后门准则和 do-演算方法进行因果推理的方法则是结构化因果模型区别于潜在结果模型的重要特征。一般来说，结构化因果模型的应用较为局限，因为结构化因果模型依赖因果图，而在处理现实中的问题时常常由于问题本身的复杂性而无法获得完整的先验因果图或无法分析变量间的结构，因此完全假设不成立，也就无法应用结构化因果模型。而潜在结果模型由于其对先验知识的要求不高，应用相对广泛。相对于潜在结果模型，结构化因果模型能够更好地反映节点因果特征。在 Pearl 的因果理论框架下，结构化因果模型的可解释性更强。

## 3 基于模型的因果推断

上述方法中，不难发现，这些方法往往仅仅满足于发掘节点间的因果关联，不考虑网络节点间的因果关系的演化。然而在某些领域，例如动物群体行为分析、传染病动力学研究等，网络因果结构的演化也受到了关注，因此，基于模型方法的因果推断也同样重要。本节主要根据因果模型的特性将因果模型分为线性模型的因果推断、非线性模型的因果推断和部分线性模型的因果推断。

### 3.1 线性模型的因果推断方法

线性模型的因果理论是一类用于推断假定解释变量与响应变量间是线性关系的数据的理论方法。线性模型是一类在日常生活中常见且结构较为简单的动力学模型。随着因果理论的兴起，线性因果模型的研究也逐渐开始。线性理论的因果推断首先开始于 Kano 和 Shimizu<sup>[51]</sup>用独立成分分析的方法对两类具有线性关系的非正态随机变量进行独立性分析，从而得到二者间的因果关系，该方法可以很好地识别马尔科夫等价类。在此之后 Shimizu 等<sup>[52]</sup>在文献[51]的方法基础上对原有的线性模型增加了噪声扰动，这一扰动可以很好地用于无向边的方向确认，

在方向确认后即可在原来的数据基础上给出因果图。Peters 等<sup>[53]</sup>在研究时间序列数据的因果关系时，提出对数据拟合自回归移动平均模型，对噪声项和解释变量作独立性分析来推断因果。该方法利用噪声的引入很好地对因果性进行了估计，但是也产生了一个新问题，当变量不服从正态分布时独立成分分析方法将会失效。针对此问题，Janzing 等<sup>[54]</sup>利用因果的不对称性提出了一种基于线性模型的系数矩阵和原因的协方差阵的独立性检验方法来确定变量间的因果关系，即线性迹方法。比较有意思的是，该方法对于确定的因果关系和不确定的因果关系同样有效，在高维数据场景下线性迹方法也表现得优于当时的其他算法。对线性迹方法的研究由此展开，Zscheischler 等<sup>[55]</sup>提出了线性迹方法的统计检验方法，该检验方法的检验统计量选取为原因变量和结果变量各自的协方差阵的迹的平均值。除此之外还证明了线性迹方法的优越抗噪性。

### 3.2 非线性模型的因果推断方法

由于线性模型无法对非线性特征进行刻画，而现实中大多数关系都是非线性的，因此就需要发展非线性因果模型。关于非线性因果模型的研究，Hoyer 等<sup>[56]</sup>基于文献[51]的线性加噪声模型设计了一种针对连续数据的非线性加噪声模型算法。该方法区别于线性加噪声模型的唯一特征即为不再是线性的，而用一个非线性函数表示。此后，针对离散型数据，Peters 等<sup>[57]</sup>还对 ANM 算法进行了推广。然而跟上文提到的线性加噪声模型一样，非线性加噪声模型也十分依赖于检验变量的正态性分布，且当噪声等于零时算法失效。Zhang 和 Hyvärinen 在文献[58]中论证了当数据的维度很高时，非线性函数的估计十分困难，算法的准确率也不能得到保证。

根据上述研究可以看出，加噪声模型虽然能够很好地对非线性特征进行描述，但却过于依赖噪声项从而降低了算法的灵敏度和鲁棒性，且面对高维数据时失效。为了解决这一问题，Janzing 等<sup>[54]</sup>提出了基于信息熵的因果推断算法，该算法准确率较高，并且可以更准确地预测无向图的边的方向。但是该算法并没有解决在高维数据下算法失效的问题，仅适用于低维度的因果网络发现，甚至于在二维以上失效。为了对高维数据进行因果发现，张浩等<sup>[59]</sup>基于互信息理论提出了一种两阶段因果推断方法，该算法时间复杂度较低，可以很好地识别马尔科夫等价类。并且由于该算法的实现仅仅依赖变量间的互信息，因此它不受维度的限制且适用于各类数据。但该算法在实际使用中需要知道先验概率分布，这一信息未必可以获取。且实际工作中原因和结果的

联合概率分布通常不会固定不变,这使得互信息不易求得。这时就需要一种方法来解决非线性模型的高度复杂性和不确定性,Chen等<sup>[60]</sup>运用再生核希尔伯特空间理论提出了基于核化迹的非线性因果推断算法,该方法应用核化处理的特殊性质将原始数据的非线性关系映射成线性关系,并结合文献[54]的线性迹算法实现了非线性因果模型的构造。

### 3.3 部分线性模型的因果推断方法

为了克服线性模型的局限性以及非线性模型可能带来的高度不确定性和复杂性的问题,Engle等<sup>[61]</sup>在研究气温与用电量之间的关系时提出了如式(9)的部分线性回归模型:

$$Y = AX + f(T) + \epsilon \quad (X, T) \perp \epsilon \quad (9)$$

式中,  $X$  和  $T$  分别对应于结果变量  $Y$  的线性和非线性解释变量;  $\epsilon$  表示随机噪声,解释变量与随机噪声条件独立。从形式上看,部分线性模型是由线性回归模型和非线性回归模型发展起来的一类模型,其既是线性模型的深度拓展模型,也是非线性模型的降维模型<sup>[62]</sup>。在模型因果推断的问题上,部分线性回归模型不但能够保留线性模型的高精度优点并拥有非线性模型的强稳健性和高灵活性等特点,此外还能够规避线性模型的误差较大的缺陷以及非线性模型的“维数祸根”问题。

## 4 基于时间序列数据的因果推断框架

本质上讲,因果网络是一种贝叶斯网络,它是基于贝叶斯方法构造用于表示和分析事物间的确定性关系的一类网络,也称因果网络图。因果网络的定义由 Pearl<sup>[13,16]</sup>通过概率的方式给出,同时引入外部干预的概念对因果作用进行了定义。涵盖了因果网络中结点、边集和有向无环图等概念。可简单地认为一个被赋予某种概率分布的有向图为贝叶斯网络。

区别于第2节介绍原本的因果推断方法更倾向于发掘节点(单元)间的因果关联,第3节主要倾向于介绍网络因果演化的模型化过程;本节主要介绍了从节点时间序列数据中提取网络因果性的方法,强调网络中的因果特性。但是严格来说,本节算法所用到的基础原理和方法在第2节和第3节都有所提及,但是这些方法考虑问题的角度不同,应用场景和对因果性的描述也不尽相同。

首先回到 Granger 给出的因果定义。

- ① 原因应该发生在结果之前。
- ② 原因应包含其他方式无法获得的原因信息。

在面对实际数据时,第一个要求很容易实现,

但由于第二个条件涉及到网络中几乎所有其他节点,因此面对大规模数据想要直接对第二个条件进行检查几乎是不可能实现的。鉴于这一局限,在实际应用的因果推断算法要么限制在只有少数变量、没有时间延迟的小规模网络中<sup>[63,64]</sup>,要么弱化第二个条件,从而降低网络推理的准确性<sup>[65]</sup>。总体来说,在因果推断领域,从时间序列数据推断大规模网络仍然是一个相对开放的问题<sup>[66,67]</sup>。

基于信息的因果关系推断方法克服了传统方法中常见的依赖模型的局限性,如格兰杰因果关系检验<sup>[2,3]</sup>。Schreiber 提出转移熵作为两个过程之间信息流或有效耦合的度量,不关心二者之间的具体关系是怎样的<sup>[68]</sup>。从  $Y$  到  $X$  的转移熵的直观含义为在已知  $X$  当前状态的情况下,观测到  $Y$  当前的状态对  $X$  未来状态估计的不确定量的减少,本质上是  $X$  的未来和  $Y$  的过去的状态在  $X$  的过去状态条件下作用的互信息<sup>[69,70]</sup>。转移熵可以将时间演化和因果耦合联系起来的这一特点适用于时序的因果演化分析且不依赖于模型,这一特点使其迅速开始在包括生物信息学、神经科学、气候学和其他学科在内的广泛学科中得到广泛应用,成为推断复杂动态过程下因果网络结构的实用工具<sup>[71,72]</sup>。然而,转移熵是专门为检测两个过程之间的信息流的方向性而引入的,面对网络推理问题时有一定局限性<sup>[73]</sup>。如果没有适当的条件,基于转移熵的推理往往会产生系统误差,一个很重要的缺陷即为转移熵无法区分直接邻居和间接邻居。而条件熵概念的提出很好地解决了这一问题,Sun 等证明了一般随机过程的最优因果熵原理,简单来说,即一个节点的因果父节点集即使为之有最大因果熵的最小节点集合。这个原理把因果推理的问题转化为因果熵的优化。并进一步通过简单的贪婪算法来解决这一组合优化问题,该算法效率较高,且准确率较高。适用于大规模网络的因果推断和因果网络的生成。这些特点使之可以分析个体行为之间的关联,挖掘群体中个体之间的关联和网络结构,从而对群体智能行为进行分析;文献[74,75]中将最优因果熵原理应用到鸽群群体行为分析中,最终得到了鸽群演化过程中的因果演化。由此我们可以看到,因果推断方法在群体智能分析中可以起到推断数据结构的重要作用。

## 5 总结

因果推理与机器学习领域有着密切的关系。近年来,机器学习领域的蓬勃发展促进了因果推理领域的发展。一些机器学习方法,如决策树、集成方



法、深度神经网络等可以准确地估计或预测潜在的结果。此外，机器学习方法也提供了一个新的处理混杂因素的思路。利用深度表征学习的生成对抗神经网络等方法，可以消除混杂变量的额外影响，从而使目的变量独立于混杂变量。除此之外，在机器学习中，数据量的大小直接影响训练结果的准确性，通常来说，数据量越大，训练结果越精确。然而，在因果推断中，数据量的大小对结果有效性的影响是十分有限的，更大的数据集可以得到更精确的估计效果，但不能确保这些估计是正确且无偏的。机器学习方法促进了因果推理的发展，同时因果推理也有助于机器学习方法的发展。对于现代机器学习的研究来说，单纯追求预测准确性是不够的，越来越多的研究者开始关心预测结果的无偏性和可解释性。当前，在强化学习和推荐系统等领域已经有了很多和因果推断相关的方法。

本文综述了因果模型和因果推断的现有理论框架和算法，但是值得一提的是，目前因果推断的相关理论和建模仍然在发展的过程中，业内仍然没有形成统一的框架和理论结构。因此自然而然可以提出一个问题：因果推断的发展和应用在未来会是什么样的呢？随着人类进入智能化时代，各种人工智能算法纷纷涌现，因果理论作为一种对现实世界可以良好描述的逻辑性语言，其与人工智能的结合应当是因果理论的重要研究方向。目前的机器学习系统几乎完全是在黑盒模型下运行的，这对机器学习的发展造成了一定限制。这样的系统不能进行推理干预，因此不能作为强大的人工智能的基础。为了实现媲美人类水平的智能，机器学习需要现实模型的指导，这一点类似于因果推理中使用的模型。为了展示这些模型的重要作用，Pearl 曾经提到过经典概念“因果革命”。在强人工智能时代<sup>[76]</sup>，人工智能可以进行反事实思考，实现真正的“智能”。而因果理论将为强人工智能时代的机器学习算法提供如下可能的思路 and 启发。

#### ① 编码因果假设（机器学习的可解释性）

结构化因果模型的图结构可以给机器学习提供一些可解释性基础，从因果图结构中可以判断反事实或统计相关性。通过 d-separation 的图形标准，可以直接对独立性进行检验。这一框架提供了因果和概率之间的基本联系。

#### ② Do 算子和混淆变量的控制

混淆变量长期以来一直被认为是从数据得出因果性的主要障碍。而“后门准则”可以很好地解决这一问题。在因果理论框架下，选择一组适当的协变量来控制混杂的任务可以很容易地实现。在机

器学习中，同样采取因果的手段去除混淆变量的影响以提升精度。

#### ③ 反事实算法化

在 Pearl 看来，反事实是强人工智能的思考，也是人类智能与机器智能的本质区别，构造反事实算法的机器学习理论是实现强人工智能的一个重要思路。

#### ④ 直接和间接效应估计

解释将变化从原因传递到结果的机制也是人工智能的重要课题，而这一解释必须援引反事实分析。反事实结构可以定义和区分直接和间接影响，将其加以区分也是人工智能的重要一步。

#### ⑤ 外部效度与样本选择偏差

机器学习的实验有效性受到实验设置的影响。训练环境改变会导致训练结果的差异。而基于因果的计算方法提供了一个克服由于训练变化而产生的偏差的完整方法，既可以用于重新调整已学习的策略，以规避环境变化，也可以用于控制非代表性样本造成的偏差，以提高机器学习理论的鲁棒性。

#### ⑥ 缺失数据

使用缺失过程的因果模型，可以形式化条件。在这些条件下，从不完整的数据中恢复因果关系和概率关系，并且满足一定条件时，估计是一致且无偏的<sup>[78]</sup>。

#### ⑦ 因果发现

因果性本身就是一种逻辑性的数据结构，可以通过因果性论述数据结构。

因果推断和因果模型的发展及其所带来的“因果革命”是否会是人类迈向更高智能文明的一种途径，这是一个有价值的课题。

## 参考文献(References)

- [1] Holland P W. Statistics and Causal Inference[J]. Journal of the American Statistical Association, 1986, 81: 945-970.
- [2] Granger C W J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods[J]. Econometrica, 1969(37): 425-438.
- [3] Granger C W J. Some Recent Developments in a Concept of Causality[J]. Econometrics, 1988(39): 199-211.
- [4] Altman N, Krzywinski M. Points of Significance: Association, Correlation and Causation[J]. Nature Methods, 2015, 12(10): 899-900.
- [5] Pearl J. Causal Inference in Statistics: an Overview[J]. Statistics Surveys, 2009, 3: 96-146.
- [6] Stephen M, Christopher W. Counterfactuals and Causal Inference: Methods and Principles for Social Research[M]. Cambridge: Cambridge University Press, 2007.
- [7] Stuart E A. Matching Methods for Causal Inference: a Review and a Look Forward[J]. Statistical Science: a Review Journal of the Institute of Mathematical Statistics, 2010, 25(1): 1.



- [8] Athey S, Imbens G W. Machine Learning Methods for Estimating Heterogeneous Causal Effects[J]. *Statistics*, 2015, 1050(5): 1-26.
- [9] Chakraborty B, Murphy S A. Dynamic Treatment Regimes[J]. *Annual Review of Statistics and Its Application*, 2014, 1: 447-464.
- [10] Schwab P, Linhardt L, Bauer S, et al. Learning Counterfactual Representations for Estimating Individual Dose-response Curves[M]. 2019, preprint arXiv: 1902.00981.
- [11] Rubin D B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies[J]. *Journal of Educational Psychology*, 1974, 66(5): 688.
- [12] Splawa N J, Dabrowska D M, Speed T. On the Application of Probability Theory to Agricultural Experiments[J]. *Statistical Science*, 1990, 5(4): 465-472.
- [13] Pearl J. Causal Diagrams for Empirical Research[J]. *Biometrika*, 1995, 82(4): 669-688.
- [14] Pearl J. Causality[M]. Cambridge: Cambridge University Press, 2009.
- [15] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference[M]. Amsterdam: Elsevier, 2014.
- [16] Pearl J. Causality: Models, Reasoning and Inference[M]. New York: Springer, 2000.
- [17] Rubin D B. Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies[J]. *Journal of Educational Psychology*, 1974, 66(5): 688-701.
- [18] Hume D. An Enquiry Concerning Human Understanding[M]. LaSalle IL: Open Court Press, 1959.
- [19] Mill J S. A System of Logic in Collected Works of John Stuart Mill[M]. Toronto: University of Oronto Press, 1973.
- [20] Neyman J. On the Application of Probability Theory to Agricultural Experiments[J]. *Statistical Science*, 1990, 5(4): 465-480.
- [21] Fisher R A. Statistical Methods for Research Workers[M]. Edinburgh: Oliver and Boyd, 1925.
- [22] Fisher R A. Design of Experiments[M]. Edinburgh: Oliver and Boyd, 1935.
- [23] Heckman J J, Vytlacil E J. Econometric Evaluation of Social Programs, Part II[M]. Amsterdam: Elsevier, 2007.
- [24] Splawa N J, Dabrowska D M, Speed T. On the Application of Probability Theory to Agricultural Experiments[J]. *Statistical Science*, 1990: 465-472.
- [25] Pearl J. Judea Pearl on Potential Outcomes[E/OL]. <http://causality.cs.ucla.edu/blog/index.php/2012/12/03/judea-pearl-on-potential-outcomes/>.
- [26] Imbens G W, Rubin D B. Causal Inference in Statistics, Social, and Biomedical Sciences[M]. Cambridge: Cambridge University Press, 2015.
- [27] Johansson F, Shalit U, Sontag D. Learning Representations for Counterfactual Inference[J]. New York: International Conference on Machine Learning, 2016.
- [28] Shalit U, Johansson F D, Sontag D. Estimating Individual Treatment Effect: Generalization Bounds and Algorithms[C]. Sydney: 34th International Conference on Machine Learning, 2017.
- [29] Pearl J. Causality: Models, Reasoning, and Inference[M]. New York: Cambridge University Press, 2009.
- [30] Goldberger A. Structural Equation Models: an Overview[M]. New York: Seminar Press, 1973.
- [31] Pearl J. Probabilistic Reasoning in Intelligent Systems[M]. San Francisco: Morgan Kaufmann Publishers, 1988.
- [32] Lauritzen S. Graphical Models[M]. Oxford: Clarendon Press, 1996.
- [33] Spirtes P, Glymour C, Scheines R. Causation, Prediction, and Search[M]. Cambridge: MIT Press, 2000.
- [34] Rosenbaum P, Rubin D. The Central Role of Propensity Score in Observational Studies for Causal Effects[J]. *Biometrika*, 1983, 70: 41-55.
- [35] Bareinboim E, Pearl J. Causal Inference and the Data-fusion Problem[J]. *Proceedings of the National Academy of Sciences*, 2016, 113(27): 7345-7352.
- [36] Li A, Pearl J. Unit Selection Based on Counterfactual Logic[J]. Macao: 28th International Joint Conference on Artificial Intelligence, 2019.
- [37] Dieter K. A Software Package for Sequential Quadratic Programming[M]. Oberpfaffenhofen: Deutsche Forschungund V Ersuchsanstalt für Luftund Raumfahrt Köln, 1988.
- [38] Greenland S, Pearl J, Robins J. Causal Diagrams for Epidemiologic Research[J]. *Epidemiology*, 1999(10): 37-48.
- [39] Glymour M, Greenland S. Causal Diagrams[J]. *Modern Epidemiology*, 2008: 183-209.
- [40] Cox D, Wermuth N. Causality: a Statistical View[J]. *International Statistical Review*, 2004, 72: 285-305.
- [41] Lauritzen S. Causal Inference from Graphical Models[M]. Boca Raton: Chapman and Hall/CRC Press, 2001: 63-107.
- [42] Morgan S, Winship C. Counterfactuals and Causal Inference: Methods and Principles for Social Research (Analytical Methods for Social Research)[M]. New York: Cambridge University Press, 2007.
- [43] Dawid A. Influence Diagrams for Causal Modelling and Inference[J]. *International Statistical Review*, 2002, 70: 161-189.
- [44] Woodward J. Making Things Happen[M]. New York: Oxford University Press, 2003.
- [45] Rothman K. Causes[J]. *American Journal of Epidemiology*, 1976, 104: 587-592.
- [46] Meek C, Glymour C. Conditioning and Intervening[J]. *British Journal of Philosophy Science*, 1994, 45: 1001-1021.
- [47] Greenland S, Pearl J, Robins J. Causal Diagrams for Epidemiologic Research[J]. *Epidemiology*, 1999, 10: 37-48.
- [48] Cole S, Hernán M. Fallibility in Estimating Direct Effects[J]. *International Journal of Epidemiology*, 2002, 31: 163-165.
- [49] Arah O. The Role of Causal Reasoning in Understanding Simpson's Paradox, Lord's Paradox, and the Suppression Effect: Covariate Selection in the Analysis of Observational Studies. *Emerging Themes in Epidemiology*[E/OL]. <http://www.eteonline.com/content/5/1/5>.
- [50] Shrier I. Letter to the Editor: Propensity Scores[J]. *Statistics in Medicine*, 2009, 28: 1317-1318.
- [51] Kano Y, Shimizu S. Causal Inference Using Nonnormality[C]. Tokyo: International Symposium on Science of Modeling: the 30th Anniversary of the Information Criterion, 2003.
- [52] Shimizu S, Hoyer P O, Hyvarinen A. A Linear Non-gaussian Acyclic Model for Causal Discovery[J]. *Journal of Machine Learning Research*, 2006, 7(10): 2003-2030.
- [53] Peters J, Janzing D, Grettton A. Detecting the Direction of Causal Time Series[C]. Montreal: 26th International Conference on Machine Learning, 2009.
- [54] Janzing D, Hoyer P O, Schölkopf B. Telling Cause from Effect Based on High-dimensional Observations[C]. Haifa: 27th International Conference on Machine Learning, 2010.
- [55] Zscheichler J, Janzing D, Zhanag K. Testing Whether Linear Equations are Causal: a Free Probability Theory Approach[J]. Montreal: 25th Conference on Uncertainty in Artificial Intelligence, 2012.
- [56] Hoyer P O, Janzing D, Mooij J. Nonlinear Causal Discovery with Additive Noise Models[C]. Cambridge: Advances in Neural Information Processing Systems, 2009.

- [57] Peters J, Janzing D, Schölkopf B. Causal Inference on Discrete Data Using Additive Noise Models[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2011, 33(12): 2436-2450.
- [58] Zhang K, Hyvärinen A. On the Identifiability of the Post-nonlinear Causal Model[C]. Stockholm: 25th Conference on Uncertainty in Artificial Intelligence, 2009.
- [59] 张浩, 郝志峰, 蔡瑞初, 等. 基于互信息的适用于高维数据的因果推断算法[J]. 计算机应用研究, 2015, 32(2): 382-385.  
Zhang H, Hao Z F, Cai R C, et al. High Dimensional Causality Discovering Based on Mutual Information[J]. Application Research of Computers, 2015, 32(2): 382-385.
- [60] Chen Z T, Zhang K, Chan L W. Nonlinear Causal Discovery for High Dimensional Data: a Kernelized Trace Method[C]. Texas: IEEE 13th International Conference on Data Mining, 2013.
- [61] Engle R F, Granger C W J, Rice J. Semiparametric Estimators of the Relation Between Weather and Electricity Sales[J]. Journal of American Statistics Association, 1986, 80: 310-319.
- [62] 周衍琪, 刘静颐, 王越. “部分线性模型”研究文献综述[J]. 商, 2015, 13: 296.  
Zhou Y Q, Liu J Y, Wang Y. Literature Review of “Partial Linear Model”[J]. Business, 2015, 13: 296.
- [63] Guo S, Seth A K, Kendrick K M, et al. Partial Granger Causality-eliminating Exogenous Inputs and Latent Variables[J]. Neuroscience Methods, 2008, 172: 79-93.
- [64] Sun J, Boltt E M. Causation Entropy Identifies Indirect Influences, Dominance of Neighbors and Anticipatory Couplings[J]. Journal of Physics D: Applied Physics, 2014, 267: 49-57.
- [65] Vicente R, Wibral M, Lindner M, et al. Transfer Entropy—a Model-free Measure of Effective Connectivity for the Neurosciences[J]. Journal of Computational Neuroscience, 2011, 30: 45-67.
- [66] Kuchaiev O, Rasajski M, Higham D J. Geometric De-noising of Protein-protein Interaction Networks[J]. PLOS Computational Biology, 2009, 5: 1000454.
- [67] Stolovitzky G, Monroe D, Califano A. Dialogue on Reverse-engineering Assessment and Methods: the DREAM of High-throughput Pathway Inference[J]. Annals of the New York Academy of Sciences, 2007, 1115: 1-22.
- [68] Schreiber T. Measuring Information Transfer[J]. Physical Review Letters, 2000, 85(2): 461-465.
- [69] Kaiser A, Schreiber T. Information Transfer in Continuous Processes[J]. Journal of Physics D: Applied Physics, 2002, 166: 43-62.
- [70] Palus M, Komárek V, Hrnčíř Z, et al. Synchronization as Adjustment of Information Rates: Detection from Bivariate Time Series[J]. Physical Review E, 2001, 63: 046211.
- [71] Boltt E. Synchronization as a Process of Sharing and Transferring Information[J]. International Journal of Bifurcation and Chaos, 2012, 22: 1250261.
- [72] Vicente R, Wibral M, Lindner M, et al. Transfer Entropy—a Model-free Measure of Effective Connectivity for the Neurosciences[J]. Journal of Computational Neuroscience, 2011, 30: 45-67.
- [73] Smirnov D A. Spurious Causalities with Transfer Entropy[J]. Physical Review E, 2013, 87: 042917.
- [74] Chen D, Kang M, Yu W. Probabilistic Causal Inference for Coordinated Movement of Pigeon Flocks[J]. EPL, 2020, 130(2): 28004.
- [75] Chen D, Wang Y, Wu G, et al. Inferring Causal Relationship in Coordinated Flight of Pigeon Flocks[J]. Chaos, 2019, 29(11): 113118.
- [76] Pearl J. Theoretical Impediments to Machine Learning with Seven Sparks from the Causal Revolution[J]. Los Angeles: 11th ACM International Conference on Web Search and Data Mining, 2018.
- [77] Mohan K, Pearl J. Graphical Models for Processing Missing Data[J]. Forthcoming, Journal of American Statistical Association, 2020, 10: R-473.