

因果机器学习在医疗决策中的应用研究综述

周婵 魏郑希 刘江 陈芋文

中国科学院重庆绿色智能技术研究院 重庆 400700

(zhouchan24@mails.ucas.ac.cn)

摘要 归纳了因果机器学习的核心概念和基本原理,以及其在医疗领域的应用研究进展,为医学研究人员、医生和决策者提供重要的参考。介绍了因果学习的基本概念、主要因果模型以及因果机器学习模型,系统梳理了因果机器学习在医疗决策中的应用进展及面临的挑战。因果机器学习相关技术能够有效应用于医疗诊断、治疗、预测等过程中,增强对病症的把控和识别能力,从而帮助医生和决策者更好地理解 and 预测治疗效果,为患者提供更加有效的医疗方案。因此,因果机器学习在医疗决策中具有广阔的应用前景,但当下仍然面临着数据质量、模型可解释性等方面的挑战。未来的研究应关注如何克服现有挑战,提供更加精确和个性化的医疗决策支持,更大程度地维护患者的健康。

关键词: 因果机器学习;医疗决策;医疗应用

中图分类号 TP391

Research Progress on Application of Causal Machine Learning in Medical Decision-making

ZHOU Chan, WEI Zhengxi, LIU Jiang and CHEN Yuwen

Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400700, China

Abstract This paper summarises the core concepts and fundamentals of causal machine learning, as well as the research progress of its application in healthcare, providing an important reference for medical researchers, doctors and policy makers. It introduces the basic concepts of causal learning, the main causal models, and causal machine learning models, systematically sorts out the progress and challenges of the application of causal machine learning in medical decision-making. This paper points out that causal machine learning-related technologies can be effectively applied to the process of medical diagnosis, treatment, and prediction to enhance the ability to control and identify the disease, thus helping doctors and decision makers better understand and predict the treatment effect, and provide more effective medical solutions for patients. Therefore, causal machine learning has a broad application prospect in medical decision-making, but it still faces challenges in data quality and model interpretability at the moment. Future research should focus on how to overcome the existing challenges, provide more precise and personalised medical decision support to maintain patients' health to a greater extent.

Keywords Causal machine learning, Medical decision-making, Medical application

1 引言

医疗决策^[1]是一个复杂的过程,需要综合考虑患者的临床症状、检查结果、治疗方案等多方面因素。医疗决策过程通常涉及收集、解释和利用大量的医疗数据、临床信息和专业知识,以做出最佳的治疗选择。传统的医疗决策方法往往依赖于医生的经验和直觉,决策结果易受到医生主观偏见的影响,这在一定程度上不仅限制了决策的客观性和系统性,也使得决策过程难以追溯和解释。随着现代医疗环境的发展,这些局限性愈发凸显,迫切需要更为科学、系统、透明的决策辅助工具和方法。

随着大数据和人工智能技术的发展,医疗决策领域开始引入先进的数据分析方法,以期提升决策的科学性和精准度。机器学习方法可识别模式并发现相关性,但无法解释做出预

测或判断的本质原因和规律^[2]。而因果机器学习利用因果推断^[3]的方法,提供了一个更精准的框架,能够理解疾病发展和治疗效果之间的内在联系,识别并量化因果关系,为医疗决策提供更加科学和可靠的依据,帮助医生进行疾病诊断、治疗选择和预后评估,提高治疗的安全性和有效性。

目前因果机器学习的相关技术已在医疗诊断、治疗和预后等决策过程中逐渐应用并展现出重要价值。例如,在疾病诊断中,它可以帮助医生理解不同症状之间的因果联系,从而更加准确地识别疾病原因;在治疗决策中,它能够预测不同治疗方案的潜在效果,为个性化治疗提供依据;在预后评估中,它能够评估疾病进展和治疗结果之间的因果路径,帮助医生制定更有针对性的干预措施。尤其在早期医疗决策阶段,因果机器学习方法为识别疾病风险、优化干预策略、预测治疗效果以及评估医疗政策等方面提供了强大的支持。

基金项目:国家自然科学基金(62371438);重庆市自然科学基金项目(CSTB2024NSCQ-MSX1043)

This work was supported by the National Science Foundation of China (62371438) and Chongqing Municipal Natural Science Foundations (CSTB2024NSCQ-MSX1043).

通信作者:陈芋文(chenyuwen@cigit.ac.cn)

因果机器学习技术已在医疗决策领域展现出巨大潜力。本文将综述因果机器学习在医疗决策中的应用,从因果机器学习的基本技术原理出发,介绍其在疾病诊断、治疗决策、预后预测等关键医疗环节中的应用,并探讨当前研究中存在的挑战和未来的发展方向。通过对这些内容的深入分析,本文旨在为医疗决策领域提供新的视角和思路,推动医疗决策向更科学、更精准的方向发展。

2 因果机器学习的基本原理

随着技术的不断发展,机器学习模型的性能也在不断提高。为了将机器学习模型应用于各种复杂的真实应用场景,机器学习领域开始借助因果分析的方法和技术来解决其面临的各种问题。因果机器学习引入了机器学习技术和因果思想的研究方向,它专注于使用统计和计算方法来推断数据中的因果关系。与机器学习模型不同,因果机器学习模型不仅试图找出数据中的模式和关联,而且还试图理解导致这些模式的潜在因果机制。因果机器学习的目标是建立能够揭示变量间因果效应的模型,这些模型可以用于做出更好的预测,生成决策支持的解释,以及进行有效的干预。因果推断是统计学的一个领域,它研究如何从观测数据中推断变量之间的因果关系。因果机器学习通常依赖于因果推断的理论基础,然后引入了机器学习的技术和算法,以处理更复杂的数据和模型。本章介绍因果理论的基本概念、主要的因果模型以及因果机器学习模型。

2.1 基本概念

因果关系对现代社会的认知起到了至关重要的作用。基于人类认知的发展,Pearl^[2]将因果关系划分为3个层级:关联(Association)、干预(Intervention)和反事实(Counterfactuals)。关联着重于识别变量间的共变趋势。然而,关联并不等同于因果联系——即所谓的“相关性不等于因果性”原则。关联分析虽然能够揭示潜在的模式和趋势,但它无法确定因果方向或排除隐藏变量的影响。干预是对研究样本施加的主动操作,是因果关系判定的重要操作。通过比较干预前后或干预组与对照组之间的差异,能够更深入地洞察因果关系。反事实是对一组已经观测到的变量,假设其中部分变量取另外一组值所得到的结果,比如“如果某个研究对象没有采取A操作,而是采取B操作,与现在得到的结果会有何不同?”反事实推理使我们能够深入理解因果链条中的关键转折点,为评估决策后果和设计前瞻性策略提供了更全面的思考。

潜在结果(Potential Outcome)指的是在给定的干预状态下,每个研究对象理论上可能展现的所有结果集合。事实结果(Observed Outcome)则是指在实际干预条件下观测到的研究结果。

效应(Effect)是评价干预措施影响的关键指标,它量化了干预与非干预状态下潜在结果的差异。通过系统比较不同干预下潜在结果的分布,能够准确评估干预措施的实际效益或负面影响。

分配机制(Assignment Mechanism)规定了干预措施如何被分配给不同的研究对象,即哪些研究对象对应采取了哪些干预。对于一个二值干预, $X=1$ 代表干预组, $X=0$ 代表对照组, Y_x 表示接受相应干预 $X=x$ 后的潜在结果。

2.2 反事实模拟模型

反事实模拟模型(Counterfactual Simulation Model,

CSM)是一种通过比较实际发生的情况与相关反事实情况下可能发生的情况来做预测的因果判断模型。它通过构建一个假设的情景,改变某些关键变量或条件,然后来评估这些改变对结果的影响。CSM假设了因果关系的不同方面,还深入探讨了原因对结果发生可能性与方式的影响作用,以及原因本身的充分性与可靠性。

Lewis^[5]给出了反事实框架的符号化表达,结合可能世界语义学和反事实来刻画因果依赖性,并形成了“界定可比较相似性→用相似性来说明反事实→用反事实来定义反事实依赖性→用反事实依赖性来阐述因果依赖性→用因果依赖性来解释因果性”的逻辑链条。

CSM的分析过程分为两个关键步骤。首先,使用一种细粒度的差异识别技术,从众多变量中筛选出所有可能构成因果关联的关键变量。然后,使用额外的反事实检验来预测每个确定的变量导致结果发生的程度。定义一个完全确定的模型 M ,已知外生变量集 U 和内生变量集 V ,外生变量是由模型外部因素决定的,它们被视为模型内部因果关系的根源,不受模型内其他变量的影响;内生变量是由模型内部因素决定的,彼此之间存在因果关系。在外生变量 U 保持不变的前提下,通过改变内生变量 X 的值来观测其对结果变量 Y 的影响,即在 $x=aU$ 的情况下计算反事实 $Y_x(u)$ 。用 $X=x$ 代替 $X=aU$ 得到模型 M_x ,再带入 $U=u$,得到 $Y_x=bx+u$ 。对于任何结构因果模型 M ,反事实可以形式化为 $Y_x(u)=Y_{M_x}(u)$ 。

CSM基于以下3个关键假设:

1)除了处理本身之外,没有其他未观测到的因素同时影响个体的结果^[6],只有产生影响的事件才是原因。

2)所有影响处理分配和结果的变量都已经被观测到并包含在模型中。不存在未观测到的变量同时影响处理分配和潜在结果,从而产生混淆效应^[7]。

3)因果判断对关键变量产生影响的不同方式很敏感。例如,这些变量可能会影响结果的发生,或它是如何发生的^[8]。

Jonathan等^[9]提出了一种新的医学诊断方法,该方法使用反事实推理来确定症状和疾病之间的因果关系。经过研究论证了他们开发的反事实诊断算法优于标准关联算法,并达到了与临床专家相当的准确性。

2.3 结构因果模型

结构因果模型(Structural Causal Model, SCM)是由图灵奖获得者、著名计算机科学家Judea Pearl教授在1995年提出的理论框架^[10],它为系统化地描述和分析复杂系统中变量间的因果关系提供了一种强有力的工具。结构因果模型结合了数学表达、图形表示以及逻辑推理,为从观测数据中推断因果效应提供了坚实的理论基础。

结构因果模型通过一组结构方程 $f=\{f_X:W_X\rightarrow X|X\in V\}$ 来定义模型中的因果关系,其中 $W_X\subseteq(U\cup V)-\{X\}$, U 是外生变量集, V 是内生变量集。这些方程描述了每个内生变量 X 如何根据其父辈变量(直接因果影响它的外生或内生变量)以及自身的外生变量(如果有的话)的值来确定其值。

每个内生变量 X_i 对应一个结构方程:

$$X_i=f_i(pa(X_i),U_i) \quad (1)$$

其中, $pa(X_i)$ 表示 X_i 的所有直接因果前驱(父节点),即影响 X_i 的其他内生变量; U_i 是与 X_i 相关的外生变量(或称未观测到的混淆变量、遗漏变量),代表了除已知父节点影响之外,影

响 X_i 的其他不可控因素或随机扰动; f_i 是一个确定性函数,描述了在给定父节点值和外生变量值时, X_i 的生成机制。

结构因果模型的核心在于其图模型,这是一种定性的表示方法。SCM 通常伴随一个有向无环图(Directed Acyclic Graph,DAG),该图直观地展示了变量间的因果关系结构,其中节点代表变量,箭头表示因果方向,即从原因指向结果。因果图不仅清晰地展现了哪些变量对其他变量具有直接的因果作用,还展现了变量间的潜在间接影响路径和可能的混杂因素,这些因素可能会对因果效应产生干扰。

对于系统的联合分布,SCM 提供了 3 种不同的表达方式。1)建立完整的概率表,为所有可能的变量取值组合都分配相应的概率值。这种方法在变量数量较少时是可行的,但在处理多变量问题时,由于需要巨大的存储空间和计算资源,变得不太实用。2)利用结构因果模型的 n 个结构方程加上误差项来计算联合分布的所有概率。这种方法在理论上是可行的,但在实践中,函数的具体形式和误差项的分布可能是未知的,或者难以准确表达的。3)图模型结合乘法分解法则。这种方法利用图模型的结构特性,通过乘法分解法则来简化联合分布的表达和计算。对于任意有向无环图模型,变量的联合分布均可通过各节点的条件分布概率之积计算出。

$$P(x_1, x_2, \dots, x_n) = \prod_i P(x_i | p a_i) \quad (2)$$

其中, $P(x_1, x_2, \dots, x_n)$ 表示变量 x 的联合分布。这种方法可以有效地处理复杂的多变量分布,尤其是在变量间存在复杂依赖关系时。

图模型的深层意义在于将一个复杂的“高维”分布估计问题转换为多个简单的“低维”分布估计问题。不再需要依赖于数据覆盖全体变量的所有取值组合来进行频次计算,而仅仅需要在每个子问题中考虑部分变量的取值组合。因此,对于每个问题而言,可用的数据量相对增加,从而显著提高了频次统计的准确性和可靠性。

Reinhold 等^[11]开发了一个结构因果模型,为多发性硬化症患者生成大脑的反事实 MR 图像,可用于对疾病进展进行建模,并用于需要控制混杂因素的下流图像处理任务。

2.4 潜在结果框架

在因果推断中,每个个体对于每一个可能的干预状态(例如,接受治疗或不接受治疗),都有一个或多个潜在的结果。这些干预可以是政策实施、医疗措施或者任何形式的活动。以医疗治疗为例,若个体 i 服用了药物治疗,那么观测到的干预状态是 $D_i=1$,可以观测到潜在结果为 Y_{1i} ;若个体 i 未接受药物治疗,那么观测到的干预状态是 $D_i=0$,对应的潜在结果为 Y_{0i} 。对于个体 i ,某项干预的因果效应为两种状态下潜在结果的比较: $\tau=Y_{1i}-Y_{0i}$ 。当干预状态实现后,仅能观测到实现状态下的潜在结果,而未实现状态下的潜在结果是不可观测的。

潜在结果模型(Potential Outcomes Framework)也称为 Neyman-Rubin 潜在结果或 Rubin 因果模型,是著名统计学家 Donald Rubin 教授在 1978 年提出的^[12],是一种用统计学语言来描述因果关系的模型^[13]。该模型通过设定对照组(未接受干预)和干预组(接受干预),考虑每个个体在所有可能的处理条件下(如接受治疗或未接受治疗)的潜在结果(如治愈或未治愈),从而定义和量化因果效应。

假设对于每个个体 i ,存在两个潜在结果 Y_{0i} 和 Y_{1i} 。而个

体 i 实际观察到的结果是这两个潜在结果之一,这取决于其实际接受的干预状态 D_i 。因果效应,例如平均因果效应(Average Causal Effect,ACE)可以通过比较两种处理状态下的潜在结果来计算:

$$ACE = E[Y_{1i} - Y_{0i}] \quad (3)$$

潜在结果框架的核心在于通过对比同一个体在不同处理状态下的反事实结果,来推断处理对结果的因果效应,而无需直接观察到所有可能的结果状态同时发生。

潜在结果模型存在 3 个假设条件^[14]。

假设 1 [Stable Unit Treatment Value Assumption, 稳定单元处理假设]。任何样本的潜在结果不会因分配给其他样本的治疗而变化,即个体之间是独立的。

假设 2 [Ignorability, 可忽略性]。给定一组观察到的环境变量 X 后,治疗分配 W 与治疗可能产生的潜在结果 $Y_i(0), Y_i(1)$ 是独立的。如果两个患者具有相同的环境变量 X ,则无论他们的潜在结果值如何,他们的治疗分配机制都应该相同,即不能根据治疗产生潜在结果的好坏来选择治疗。

$$p(W|X=x, Y_i(0), Y_i(1)) = p(W|X=x, Y_j(0), Y_j(1)) \quad (4)$$

如果两个患者具有相同的背景变量 X ,无论治疗分配如何,他们的潜在结果都应该相同,即治疗的分配不会影响到潜在的结果分布。

$$p(Y_i(0), Y_i(1) | X=x, W=W_i) = p(Y_j(0), Y_j(1) | X=x, W=W_j) \quad (5)$$

假设 3 [Positivity, 正向假设]。对于任何一个个体都有非零的可能性来采取每种处理方式。

$$P(W=\omega | X=x) > 0, \forall \omega \text{ and } x \quad (6)$$

2.5 因果机器学习模型

近年来,随着机器学习在医疗领域的广泛应用,许多以往由人类主导的决策过程正在逐步被自动化算法所替代。然而,许多机器学习模型因其不透明的内部机制,很难向用户提供其决策的解释。为了确保这些决策的可靠性,理解其背后的决策过程,因果模型在医疗决策领域的应用得到了迅速发展,如利用因果模型评估不同患者群体或不同医疗环境中认知障碍临床预测模型的可移植性^[15],通过因果建模方法增强基于机器学习算法识别的生物标志物的鲁棒性和泛化能力^[16],利用因果机器学习技术来提升医疗诊断的准确性^[17]等。

因果可解释模型^[18]是一种旨在增强决策过程透明度的因果机器学习模型,它通过模拟不同假设情境下的决策行为,为模型的决策提供详尽的因果解释。这类模型的关键优势在于能够揭示数据中隐藏的因果关系,并将这些复杂的关系以人类可理解的方式展现出来,从而增强用户对模型行为的信任,并促进用户对决策深层原因的理解。近年来,关于模型可解释性的定义和重要性,学术界已经提出了多种观点。Kim 等^[19]提出,一个方法具有可解释性是指该方法的结果可以被使用者正确并且高效地预测。Doshi-Velez 等^[20]将可解释性定义为机器学习模型以人类可理解的方式提供解释的能力。Gilpin 等^[21]更进一步,将可解释性定义为机器学习中能够理解和解释模型预测结果或决策过程的能力。他们指出,可解释的模型是那些能够总结神经网络行为原因、获得用户信任或者深入了解决策原因的模型,但即使是可解释的模型也可能无法准确地描述系统的操作。

传统的可解释性框架大多聚焦于分析数据中变量间的相关性,无法提供深入的解释。这一局限性源于,这些框架缺乏对特征或模型组件的变化是如何影响模型预测或样本标签的动态评估能力。例如,在医疗领域,临床医生可能需要回答类似“如果患者接受治疗 X ,该疾病将如何发展”或者“如果该患者接受治疗 Y ,是否仍然会出现结果 Z ”的问题。这类问题超越了简单的关联分析,需要应用因果关系的方法来理解干预措施对特定个体可能产生的具体影响。

在可解释模型的研究方向上, Kim 等^[22]提出了一个从观测数据中学习可解释模型的框架,研究者通过将任意监督算法转换为一个可解释的个体治疗效应估计框架,来弥合因果与可解释模型之间的差距。目前,已经提出了许多学习解释模型的算法,包括决策树^[23]、稀疏线性模型^[24]、广义加性模型^[25]、规则列表^[26]、决策集^[27]以及程序^[28-31],这些算法都可以有效地增强模型的可信度与清晰度,提高使用者对模型的理解。

因果决策树^[32] (Causal Decision Tree, CDT) 是一种结合了传统决策树的高效分治策略和潜在结果模型的因果机器学习模型,致力于发现数据中的因果关系。因果决策树的结点包含因果解释,其中每一个非叶子节点都代表了一个具有因果关系的特征,边表示该特征的值的分配,叶节点表示输出变量的值的分配。树中的每个节点都基于分层数据集进行估计,以确保个体在每个分层中只是否接受处理这一个不同的属性。它通过使用统计测试(如 Mantel-Haenszel 测试)来评估变量间的偏关联,从而识别出在特定上下文中成立的因果效应。此外,CDT 能够处理大规模数据集,具有自动化和可扩展性,是大数据分析中发现因果信号的有力工具。

因果贝叶斯网络(Causal Bayesian Networks)是一种概率图模型,它通过有向无环图来表示变量之间的因果关系。在因果贝叶斯网络中,每个节点代表一个随机变量且都有与之相关的条件概率表,描述了在给定其父节点值的情况下,节点的后验概率分布。节点之间的有向边表示因果效应的方向。因果贝叶斯网络不仅能够描述变量之间的联合概率分布,还能捕捉变量之间的因果结构,从而允许对干预效果进行预测。例如,通过改变某个变量的值来预测另一个变量的变化,以及进行反事实推理,即在假设某个变量的值不同于其实际值时,其他变量可能的值。

3 因果机器学习在医疗决策中的应用

近年来,因果机器学习在医疗决策中的应用取得了显著进展。因果机器学习具有优秀的可解释性,在医疗实践中,会为医生提供更全面的信息辅助决策。同时,因果机器学习有利于医生进行更好的个性化治疗。下面从诊断决策、治疗决策和预后预测这 3 个领域介绍因果机器学习在医疗领域中的应用。

3.1 诊断决策

因果机器学习在疾病诊断领域的应用中,通过构建精细化的疾病诊断模型,将疾病状态、患者症状与体征、医疗检查结果等关键变量纳入统一的因果框架中。在模型构建过程中,综合考量遗传背景、环境因素、生活方式和病史等多种因素对疾病发生与发展的影响,从而形成对疾病复杂病因机制的全面理解。这种方法有助于减少混淆因素的干扰,提高了

对罕见病、复杂疾病及疾病亚型的识别能力。基于因果模型的诊断结果,不仅能够提供疾病存在的概率或类别,更重要的是,它能够清晰地解释诊断的依据。

因果机器学习可以采用潜在结果框架,通过设想个体在接受不同诊断方法(如不同类型的检查、诊断流程或诊断工具)时可能的潜在诊断结果,来量化和比较各种诊断方法的因果效应。利用这一框架,全面评估不同诊断方法在准确性(如敏感性、特异性、阳性预测值等)、时效性(如诊断周转时间)、患者负担(如侵入性、不适感、心理压力)及医疗资源消耗(如检查费用、人力成本)等方面的因果效应。这种全面的评估为优化多目标诊断策略提供了科学依据,帮助医生和医疗决策者明确在特定临床情景下哪种诊断方案能够带来最优的诊断效果,同时兼顾患者偏好、医疗资源可用性与伦理考量等因素。北京大学王亦川课题组提出了因果隐马尔可夫模型^[33] (Causal Hidden Markov Model),从观测数据中分离出与疾病相关的隐藏变量,并基于 VAE 的变分框架结构进行学习,用于时间序列疾病的预测,提高预测的准确性和鲁棒性。Gerstenberg 等^[34]引入了反事实模拟模型(CSM),通过比较实际发生的情况与相关反事实情况下可能发生的情况来预测物理环境中的因果判断。Jackson 等^[35]开发了一种结合因果知识的随机森林方法,以寻找最有可能患有进展性多发性硬化症患者的共同特征。Flauzino 等^[36]利用多层感知机(MLP)方法发现了预测 MS 患者中更高扩展残疾状态量表(EDSS)的免疫特征。Bica 等^[37]使用观测数据来训练机器学习模型,并深入分析了预测治疗效果与监督学习之间的本质区别。

3.2 治疗决策

因果机器学习在构建治疗方案与预后结果之间的因果模型方面超越了传统的相关性分析。基于因果模型,医生和研究人员能够前瞻性地模拟和评估不同治疗方案在相同患者群体或特定个体上的疗效差异,包括直接效果、长期效应、副作用风险以及可能的交互作用等。Wilkinson 等^[38]指出,精准医学的目标并不仅仅是完美预测疾病,而是能够从患者的观测数据中提取出“可操作”的信心,以指导临床干预决策。因果机器学习可以帮助揭示疾病的潜在机制、评估治疗效果并支持个性化医疗。

然而,治疗效应的准确估计十分具有挑战性。在实际应用中,通常只能观察到实施的治疗结果,而无法得知未实施的替代治疗的反事实结果。为了得到缺失的反事实结果,现有的基于观测数据的因果推断方法可归纳为以下几个主要类别:重新加权方法^[39-40]、基于树的方法^[41-43]、匹配方法^[44-46]和双重稳健学习器^[47-48]。一般而言,匹配方法侧重于基于倾向得分或欧氏距离等度量,找到可比较的配对。重新加权方法则根据协变量为种群中的每个单位分配权重,使得处理组和对照组在关键协变量上达到平衡。基于树的方法利用基于树的机器学习模型,如决策树或随机森林,推导反事实结果。双重稳健学习器是另一种近期发展起来的方法,它将倾向得分加权与回归结果相结合,以产生无偏且稳健的估计量。

尽管这些方法在确定性干预的效应估计上取得了进展,但它们难以处理动态和随机干预^[49-50]。随机干预所干预的变量不再是一个确定性的变量,而是一个随概率变化的随机变量。传统的因果推断方法无法充分捕捉到治疗变量上的随

机干预,而因果推断干预估计提供了更细粒度的治疗效应估计,通过逐步量化随机干预对结果的影响,为治疗决策提供了更加精确的数据支持。

针对上述挑战,Duong 等^[51]基于非参数影响函数开发了一个随机干预效应估计器(SIE),以估计随机干预下的治疗效果。他们还构建了一个定制强化学习算法,该算法基于随机搜索求解器,寻找产生最大预期结果的最优策略,推动了因果推断研究的进展。Shen 等^[52]将因果发现方法应用于阿尔茨海默氏症的病理生理学研究,展示了因果发现方法在揭示复杂疾病中因果关系方面的潜力。研究表明,因果发现算法在发现因果结构方面优于结构方程模型(Structural Equation Modeling,SEM)。在现实世界的数据分析中,数据质量会影响所发现结构的正确性。结合先验知识和使用纵向数据有助于减少算法的潜在错误,提高发现结构的准确性。Zhang 等^[53]利用遗传相关性分析等手段,强调了导致神经质与重度抑郁症之间因果关系的基因。因此,因果发现在医疗领域对于治疗效果评估、疾病预测、风险预估、临床决策支持以及医疗资源优化等方面具有重要意义。

3.3 预后预测

因果机器学习在预后预测领域的应用,通过其对疾病发展过程的深度解析和因果效应的严谨建模,提高了临床决策的洞察力和预见性。

基于因果模型,临床医师和研究人员能够精准预测患者个体或特定疾病群体的疾病进展轨迹,包括病情恶化速度、并发症发生概率、疾病转归(如痊愈、稳定、恶化或死亡)等关键预后指标。因果模型不仅适用于静态的一次性预后预测,还能够随着患者病情变化、治疗干预或环境因素变动,实

时更新预测结果,动态调整预后风险评估。这种动态性使医生能够及时调整治疗策略,提前采取干预措施,为患者提供个体化的随访计划与健康指导,显著提升疾病管理的前瞻性和主动性。

在预后预测中,除了追求预测准确性,因果机器学习尤为重视模型的因果解释能力。因为只有揭示疾病发展的因果机制,预测结果才更具科学性和可信度,有助于医生理解预测背后的生物学逻辑,而不仅仅依赖于统计相关性所形成的模糊预测。因果机器学习模型通过因果推断方法,如反事实推理、结构因果模型参数估计等,对模型中的因果路径和效应大小进行量化估计和假设检验。这一过程不仅验证了模型的内在一致性,还确保了模型预测结果与已知医学知识和生物学原理的一致性,从而增强了预测结果的说服力。

Hartmann 等^[54]通过将贝叶斯网络整合数据和因果知识应用于 MS 的风险因素评估,以可视化依赖变量和潜在混杂因素之间的相互作用。多发性硬化症(MS)是一种慢性炎症性神经系统疾病,其患病率在过去几十年中逐渐增加^[55],该方法为理解 MS 的复杂病因提供了新的视角。

观察性研究能够为干预研究提供真实世界的人群和环境数据支持,并有助于筛选出值得进一步实验验证的干预措施。Dahabreh 等^[56]提出了一个观察性研究框架,旨在根据 6 个核心问题提供有关干预因果效应的证据:什么是因果问题?如果知道,什么数量可以回答因果问题?研究设计是什么?正在做出哪些因果假设?在原则上和实践中,如何利用观测到的数据来回答因果问题?对分析的因果解释是否站得住脚?

因果机器学习在医疗决策中的应用如表 1 所列。

表 1 因果机器学习在医疗决策中的应用

Table 1 Application of causal machine learning in medical decision-making

模型	内容	
诊断决策	因果可解释模型 ^[23]	通过模拟不同替代情况下的决策行为解释其自身的决策结果
	因果隐马尔可夫模型	通过引入隐藏变量来生成每个时间步骤的医学数据,避免学习到虚假关联特征,从而提高对其他数据分布的泛化能力
	反事实推断算法 ^[4]	通过假设与实际情况不同的条件来推断可能的结果或原因
	可解释框架 ^[27]	通过将任何监督算法转换为一个可解释的个体治疗效应估计框架,来弥合因果与可解释模型间的差距
治疗决策	随机干预效应估计器(SIE) ^[60]	估计随机干预下的治疗效果,构建了一个基于随机搜索求解器的定制强化学习算法,寻找产生最大预期结果的最优策略
	因果发现结构 ^[61]	包括基于约束、基于分数、基于混合的方法,通过搜索数据中统计依赖关系、评估不同因果结构分数来构建因果结构
预后预测	遗传相关性分析 ^[62]	利用遗传信息来评估变量之间的关联性,有 LDSC、遗传混合映射、多变量遗传分析等方法
	因果贝叶斯网络 ^[65]	使用贝叶斯网络将多个与多发性硬化症风险相关的因素纳入模型
	因果随机森林 ^[43]	应用随机森林模型来分类多发性硬化症(MS)的不同临床病程

4 挑战与展望

因果机器学习结合了机器学习技术和因果推断方法,旨在从数据中发现因果关系,而不仅仅停留在关联性分析。因果机器学习相较于因果推理与机器学习,不仅加速了因果推理的技术革新,为处理复杂场景中的混杂因素提供了创新思路,还丰富了机器学习的内涵,帮助构建更可靠、更公平的学习模型,提升了其预测和决策的质量。

因果机器学习具备以下显著优势:

1)因果解释性:因果机器学习方法旨在揭示变量间的因果关系,而不仅仅是相关关系。另外,因果模型能够进行反事实推理,这对于评估个体或群体在不同干预条件下的潜在结

果十分具有价值。

2)改善诊断准确性:因果机器学习通过模拟干预并计算因果效应,可以准确预测特定干预对结果的影响,然后指导最优干预策略的选择,如个性化治疗方案。

3)增强模型泛化能力:因果模型捕捉的是变量间的因果结构,因此能更好地应对数据分布的变动(如样本选择偏误、数据收集方式变化、外部环境变迁等),从而提高模型在不同环境(如不同医院或患者人群)中的泛化能力。因果机器学习辅助诊断技术能够利用丰富的患者数据,提供更精确和个性化的诊断,从而有望革新医疗保健行业。

4.1 挑战

尽管因果机器学习在医疗决策中具有广阔的应用前景,

但在实际应用过程中仍面临着诸多挑战。

1)数据要求高:因果推断通常需要大量的数据来建立准确的因果模型,而医疗数据通常存在缺失值、偏差等问题,难以满足因果分析的数据需求,也影响了评估因果关系的能力。

2)因果关系的复杂性:医疗领域的因果关系往往涉及多个生物标志物、环境因素和生活方式等众多因素。

3)模型的可解释性:医生和患者需要理解模型的预测依据,这要求因果机器学习模型不仅要准确,还要易于解释。

4)临床实践中的应用障碍:因果模型的验证通常涉及到对未观察到的反事实结果的推断,这使得在将因果机器学习方法真正应用于临床实践时,对模型的验证和推广存在挑战。此外,应用因果机器学习进行医疗决策,可能涉及数据隐私保护、医疗责任划分、算法公平性等复杂的法律与伦理问题。这需要与政策制定者、伦理专家等多方合作,明确相关法规与伦理准则,确保技术应用的合法性、合规性和伦理性。

4.2 展望

因果机器学习在医疗决策中的应用已经展现出显著的价值。它通过揭示数据背后的因果机制,为精准医疗、个性化治疗、疗效评估及医疗政策制定提供了有力支持。随着技术的不断进步,数据的日益积累以及因果推理理论的持续深化,因果机器学习在医疗决策中的应用前景广阔,有望成为医疗领域的重要工具。未来,以下几个关键方向的发展将进一步推动因果机器学习在医疗决策中的应用。

1)提升数据质量:持续研发与优化数据清洗及预处理技术,减少缺失值和偏差等问题对因果分析的影响,提升现有医疗数据的可用性;设计和实施更多的前瞻性队列研究和随机对照实验(RCT),生成高质量的因果证据,以弥补观察性数据的局限性。

2)优化模型设计:研发和优化更稳健的因果识别算法,提升现有因果推断方法的计算效率和鲁棒性;发展针对医疗数据特点的时间序列因果模型,捕捉疾病进展、治疗响应的动态,设计和应用动态干预策略,实现个体化和实时的医疗决策支持。

3)更好的结合临床实践:开发能够提供因果路径解释、局部效应分析及敏感性检验等功能的工具,增强模型的可解释性与透明度;设计互动式界面,使患者能够理解模型预测结果及其背后的因果逻辑,从而参与到医疗决策过程中,提升治疗的依从性和满意度。

围绕这些方向进行努力,将有助于加速因果机器学习未来在医疗领域的广泛应用,为实现精准医疗、优化资源配置、提升患者预后等方面发挥更大作用。

结束语 因果机器学习在诊断决策、治疗决策和预后预测等方面发挥了至关重要的作用。在诊断环节,因果机器学习能够帮助医生精准识别疾病类型、揭示疾病发生原因,为早期诊断和个性化诊疗提供科学依据。在治疗领域,因果机器学习技术通过评估不同治疗干预的因果效应,辅助医生制定最合适的个性化治疗方案,有助于提升治疗效果并减少不良反应。在预后预测方面,因果模型能够准确预测疾病的发展趋势,评估不同治疗方案对患者预后的影响,为临床决策提供前瞻性的信息。

因果机器学习为医疗决策提供了基于因果关系的科学且

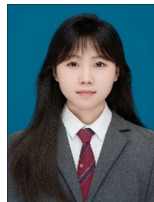
可靠的依据。凭借在揭示疾病机制、指导个性化治疗、精准预后预测等方面的优势,因果机器学习有望大幅提高医疗决策的质量,优化患者治疗效果,降低医疗成本,提升医疗服务的整体水平。尽管目前因果机器学习在医疗决策中的应用还面临着诸多挑战,但随着相关研究的不断深入、方法的持续优化以及行业生态的逐步适应,因果机器学习无疑将成为未来医疗行业中不可或缺的工具。它将对提升医疗决策的科学性和有效性产生深远的影响,推动医疗健康领域的持续进步和创新。

参考文献

- [1] CUI L, HOU Y F, ZHANG H. On medical decision support system[J]. Journal of Medical Informatics, 2003(1):28-30, 40.
- [2] PEARL J. Causal inference in statistics: An overview[J]. Statistics Surveys. 2009, 3:96-146.
- [3] PEARL J. Causal inference[C]// Proceedings of the 2008th International Conference on Causality: Objectives and Assessment. JMLR. org, 2008.
- [4] NEUBERG L G. Causality: models, reasoning, and inference, by Judea Pearl, Cambridge University Press, 2000[J]. Econometric Theory, 2003, 19(4):675-685.
- [5] LEWIS D. Causation[J]. Journal of Philosophy, 1973, 70(17):556-567.
- [6] WOODWARD J. Making Things Happen: A Theory of Causal Explanation[M]. Oxford University Press, 2004.
- [7] DANKS D. Singular Causation[M]// The Oxford Handbook of Causal Reasoning, Waldmann M R. Oxford University Press, 2017:201-216.
- [8] GLYMOUR C, DANKS D, GLYMOUR B, et al. Actual causation: a stone soup essay[J]. Synthese, 2010, 175(2):169-192.
- [9] RICHENS J G, LEE C M, JOHRI S. Improving the accuracy of medical diagnosis with causal machine learning[J]. Nature Communications. 2020, 11:3923.
- [10] PEARL J. Causal diagrams for empirical research[J]. Biometrika, 1995, 82(4):669-688.
- [11] REINHOLD J C, CARASS A, PRINCE J L. A Structural Causal Model for MR Images of Multiple Sclerosis[J]. arXiv: 2103.03158, 2021.
- [12] RUBIN D B. Bayesian Inference for Causal Effects: The Role of Randomization[J]. The Annals of Statistics. 1978, 6(1):34-58, 25.
- [13] NOGUEIRA A R, PUGNANA A, RUGGIERI S, et al. Methods and tools for causal discovery and causal inference[J]. WIREs Data Mining and Knowledge Discovery, 2022, 12(2):e1449.
- [14] YAO L, CHU Z, LI S, et al. A Survey on Causal Inference[J]. ACM Transactions on Knowledge Discovery from Data, 2021, 15(5):1-46.
- [15] FEHR J, PICCININNI M, KURTH T, et al. Assessing the transportability of clinical prediction models for cognitive impairment using causal models[J]. BMC Medical Research Methodology. 2023, 23(1):187.
- [16] PAVLOVIĆ M, AL HAJJ G S, KANDURI C, et al. Improving generalization of machine learning-identified biomarkers using causal modelling with examples from immune receptor diagnos-

- tics[J]. *Nature Machine Intelligence*, 2024, 6(1):15-24.
- [17] RICHENS J G, LEE C M, JOHRI S. Improving the accuracy of medical diagnosis with causal machine learning[J]. *Nature Communications*, 2020, 11(1):3923.
- [18] SCHÖLKOPF B. Causality for Machine Learning[M]// *Probabilistic and Causal Inference: The Works of Judea Pearl*. Association for Computing Machinery, 2022:765-804.
- [19] KIM B, KHANNA R, KOYEJO O. Examples are not enough, learn to criticize! criticism for interpretability[C]// *Proceedings of the 30th International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2016.
- [20] DOSHI-VELEZ F, KIM B. Towards A Rigorous Science of Interpretable Machine Learning[J]. arXiv:1702.08608, 2017.
- [21] GILPIN L, BAU D, YUAN B Z, et al. Explaining Explanations: An Overview of Interpretability of Machine Learning[C]// *2018 IEEE 5th International Conference on Data Science and Advanced Analytics(DSAA)*. 2018.
- [22] KIM C, BASTANI O. Learning Interpretable Models with Causal Guarantees[J]. arXiv:1901.08576, 2019.
- [23] KRZYWINSKI M, ALTMAN N. Classification and regression trees[J]. *Nature Methods*, 2017, 14(8):757-758.
- [24] TIBSHIRANI R. Regression Shrinkage and Selection Via the Lasso[J]. *Journal of the Royal Statistical Society: Series B (Methodological)*, 2018, 58(1):267-288.
- [25] USTUN B, RUDIN C. Supersparse linear integer models for optimized medical scoring systems[J]. *Machine Learning*, 2016, 102(3):349-391.
- [26] CARUANA R, LOU Y, GEHRKE J, et al. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission[C]// *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2015.
- [27] ANGELINO E, LARUS-STONE N, ALABI D, et al. Learning Certifiably Optimal Rule Lists[C]// *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Association for Computing Machinery, 2017.
- [28] ELLIS K, SOLAR-LEZAMA A, TENENBAUM J B. Unsupervised learning by program synthesis[C]// *Proceedings of the 28th International Conference on Neural Information Processing Systems*. MIT, 2015.
- [29] VERMA A, MURALI V, SINGH R, et al. Programmatically Interpretable Reinforcement Learning [J]. arXiv: 1804. 02477, 2018.
- [30] VALKOV L, CHAUDHARI D, SRIVASTAVA A, et al. HOUNDI: Lifelong Learning as Program Synthesis[C]// *Neural Information Processing Systems*. 2018.
- [31] ELLIS K, RITCHIE D, SOLAR-LEZAMA A, et al. Learning to infer graphics programs from hand-drawn images[C]// *Proceedings of the 32nd International Conference on Neural Information Processing Systems*. Curran Associates Inc., 2018.
- [32] LI Y, MA S, LE T, et al. Causal Decision Trees [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2017, 29(2):257-271.
- [33] LI J, WU B, SUN X, et al. Causal Hidden Markov Model for Time Series Disease Forecasting[C]// *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2021:12100-12109.
- [34] GERSTENBERG T, GOODMAN N D, LAGNADO D A, et al. A counterfactual simulation model of causal judgments for physical events[J]. *Psychological Review*, 2021, 128(5):935-975.
- [35] JACKSON C. Random forest classification of multiple sclerosis clinical courses[J]. *Frontiers in Neurology*, 2017, 8:1-10.
- [36] FLAUZINO T, SIMÃO A N C, DE CARVALHO JENNINGS PEREIRA W L, et al. Disability in multiple sclerosis is associated with age and inflammatory, metabolic and oxidative/nitrosative stress biomarkers: results of multivariate and machine learning procedures [J]. *Metabolic Brain Disease*, 2019, 34:1401-1413.
- [37] BICA I, ALAA A, LAMBERT C, et al. From Real-World Patient Data to Individualized Treatment Effects Using Machine Learning: Current and Future Methods to Address Underlying Challenges[J]. *Clinical Pharmacology & Therapeutics*, 2020, 109(1):87-100.
- [38] WILKINSON J, ARNOLD K F, MURRAY E J, et al. Time to reality check the promises of machine learning-powered precision medicine[J]. *The Lancet Digital health*, 2020, 2(12):e677-e680.
- [39] GRUBER S, LAAN M V D. tml: An R Package for Targeted Maximum Likelihood Estimation[J]. *Journal of Statistical Software*, 2012, 51(13):1-35.
- [40] AUSTIN P C, STUART E A. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies[J]. *Statistics in Medicine*, 2015, 34:3661-3679.
- [41] CHIPMAN H A, GEORGE E I, MCCULLOCH R E. BART: Bayesian Additive Regression Trees[J]. *The Annals of Applied Statistics*, 2008, 4:266-298.
- [42] HILL J L. Bayesian Nonparametric Modeling for Causal Inference[J]. *Journal of Computational and Graphical Statistics*, 2011, 20:217-240.
- [43] WAGER S, ATHEY S. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests[J]. *Journal of the American Statistical Association*, 2015, 113:1228-1242.
- [44] ROSENBAUM P R, RUBIN D B. The central role of the propensity score in observational studies for causal effects[J]. *Biometrika*, 1983, 70:41-55.
- [45] DEHEJIA R H, WAHBA S. Propensity Score-Matching Methods for Nonexperimental Causal Studies[J]. *The Review of Economics and Statistics*, 2002, 84(1):151-161.
- [46] HO D, IMAI K, KING G, et al. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference[J]. *Journal of Statistical Software*, 2011, 42(8):1-28.
- [47] OPRESCU M, SYRGKANIS V, BATTOCCHI K, et al. EconML: A Machine Learning Library for Estimating Heterogeneous Treatment Effects [EB/OL]. https://mirunaoprescu.com/assets/publications/2019_econml/EconML_Workshop_Poster.pdf.
- [48] DUDÍK M, LANGFORD J, LI L. Doubly Robust Policy Evalua-

- tion and Learning [C] // International Conference on Machine Learning, 2011.
- [49] DUDÍK M, ERHAN D, LANGFORD J, et al. Doubly Robust Policy Evaluation and Optimization [J]. arXiv: 1503. 02834, 2015.
- [50] TIAN J. Identifying Dynamic Sequential Plans [C] // Conference on Uncertainty in Artificial Intelligence, 2008.
- [51] DUONG T D, LI Q, XU G. Stochastic intervention for causal inference via reinforcement learning [J]. Neurocomputing, 2022, 482: 40-49.
- [52] SHEN X, MA S, VEMURI P, et al. Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology [J]. Scientific Reports, 2020, 10 (1): 2975.
- [53] ZHANG F, BARANOVA A, ZHOU C, et al. Causal influences of neuroticism on mental health and cardiovascular disease [J]. Human Genetics, 2021, 140(9): 1267-1281.
- [54] HARTMANN M, FENTON N, DOBSON R. Current review and next steps for artificial intelligence in multiple sclerosis risk research [J]. Computers in Biology and Medicine, 2021, 132: 104337.
- [55] TOPUZ K, ZENGUL F D, DAĞ A, et al. Predicting graft survival among kidney transplant recipients: A Bayesian decision support model [J]. Decision Support Systems, 2018, 106: 97-109.
- [56] DAHABREH I J, BIBBINS-DOMINGO K. Causal Inference About the Effects of Interventions From Observational Studies in Medical Journals [J]. JAMA, 2024, 331(21): 1845-1853.



ZHOU Chan, born in 2001, postgraduate. Her main research interest is medical artificial intelligence.



CHEN Yuwen, born in 1985, Ph.D. researcher, is a member of CCF (No. 38633M). His main research interests include medical and health machine learning, causal reasoning and deep learning.