

个性化反事实框架：从可穿戴设备数据生成潜在结果

Ajan Subramanian^{1*} and Amir M. Rahmani^{1,2}

摘要—可穿戴传感器数据提供了个性化健康监测的机会,但其复杂、纵向的数据流中获取可操作的见解颇具挑战性。本文介绍了一个框架,用于从多变量可穿戴设备数据中学习个性化的反事实模型。这使我们能够探索“如果”场景,以理解生活方式选择对个人特定结果的影响。我们的方法首先通过多模态相似性分析将个体数据集与类似患者的数据显示增。然后,我们使用时间 PC (Peter-Clark) 算法的适应版本来发现预测关系,建模变量在时间 $t-1$ 如何影响时间 t 的生理变化。基于这些发现的关系训练梯度提升机以量化个人特定效应。这些模型驱动反事实引擎,在假设干预下(例如活动或睡眠改变)模拟生理轨迹。我们通过一步先验预测验证和评估干预措施的合理性和影响来评价该框架。评估显示了合理的预测准确性(如平均心率 MAE 4.71 次/分钟)以及高度反事实合理性(中位数 0.9643)。关键的是,这些干预突显了个体对假设生活方式改变反应的显著个体间差异,展示了框架在个性化见解方面的潜力。这项工作提供了一种工具,用于探索个性化健康动态并生成关于个人对生活方式变化响应的假设。

Index Terms—可穿戴传感器,个性化医疗,因果发现,反事实结果,生理变化。

I. 介绍

可穿戴技术能够持续、个性化地监测生理和行为数据,为与 4P 医学(预测性、预防性、个性化、参与性)相一致的主动健康管理提供了巨大的潜力 [1]–[3]。这些设备可以跟踪心率变异性 (HRV)、睡眠、准备情况和活动等参数,提供丰富的早期检测和个人化干预的数据 [4], [5]。然而,这种数据量大、多模态且纵向的特性带来了分析上的挑战。生成具有行动价值的个体层面见解需要能够捕捉到超越简单描述性统计的个性化健康动态模型 [6]。

尽管现有研究利用可穿戴数据进行预测任务 [7]–[10],许多方法仅识别关联性而未阐明潜在机制,

可能无法充分解决个体间的差异。有效的个性化医疗需要超越通用模型,理解个体特定的反应以及睡眠、劳损和自主功能等因素之间的相互作用 [11], [12]。关键的是,虽然相关模型可以预测未来可能出现的状态,但它们往往未能探索个人如何对行为或环境的具体变化作出独特反应。理解这一点对于真正具有行动力和个性化的指导至关重要。

为了解决这一问题,我们提出了一种个性化反事实建模的框架。反事实推理,或在假设情景下对潜在结果的估计,提供了一种强大的方法来超越观察到的相关性。它们允许探究个体特定的因果关系,这对于定制干预措施至关重要,即使这些确切的情景之前并未由该个体经历过也是如此。我们的方法通过患者相似性分析增强个人数据集,以改善模型在每个人有限数据下的鲁棒性。然后我们采用一种基于约束的时间适应算法来从观察到的时间序列中估计合理的预测结构,重点关注变量如何影响个体生理状态 [13] 的日常变化 (ΔV)。这些个性化模型构成了反事实生成的基础,使我们可以提出有针对性的“如果……会怎样?”问题(例如,“如果这个人在昨晚睡眠质量很差,今天他们的活动得分可能是多少如果?”)。通过生成假设情景,我们基于学习到的个体特定动态来估计干预措施的潜在影响。

本文的关键贡献是:

- 1) 一种利用时间 PC 算法和变量变化回归 (ΔV) 从多变量可穿戴设备数据中学习个性化反事实模型的方法论。
- 2) 将患者相似性分析整合到增强个体数据集中,提高模型稳定性。
- 3) 利用这些个性化模型来估计假设干预潜在结果的反事实生成框架。
- 4) 一种全面的评估方法,既评估一步预测准确性,也评估干预措施的合理性和影响。

¹Dept. of Computer Science, University of California, Irvine. ²School of Nursing, University of California, Irvine. (*correspondence email: ajans1@uci.edu)

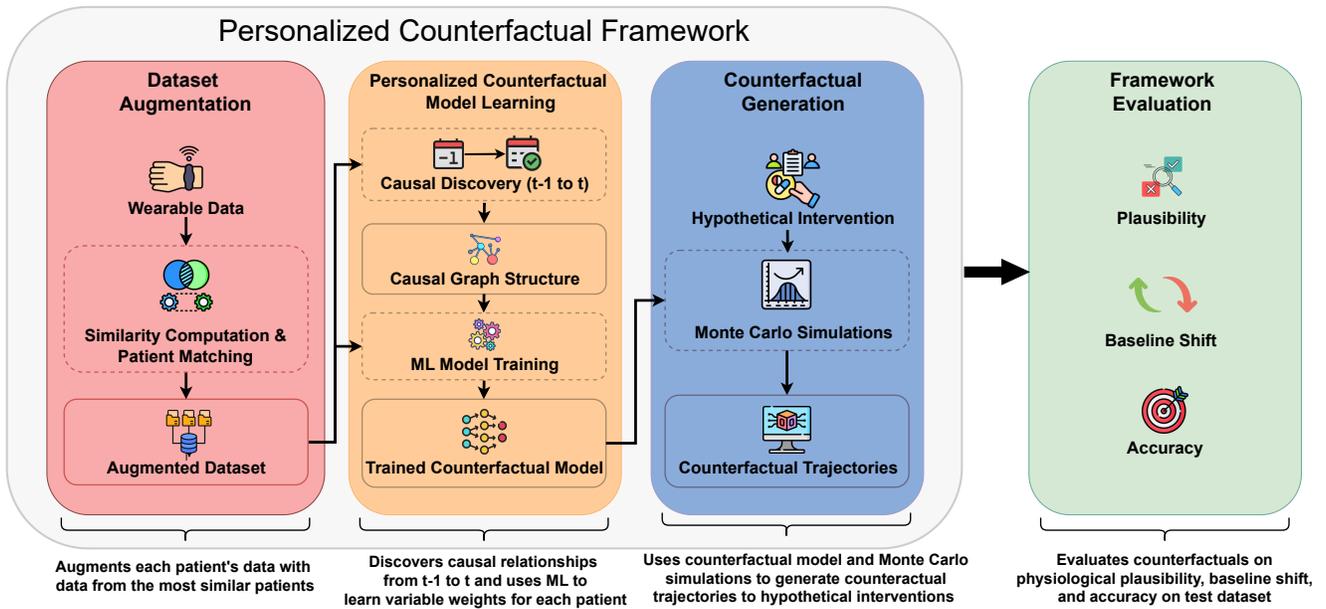


图 1: 个性化反事实框架概述。

此框架优先考虑理解个体特定的健康反应及干预措施的比较效果，而不仅仅是追求最大的预测准确性。它旨在作为一种工具，帮助生成关于个性化健康动态的新假设。

II. 方法

我们的框架（图 1）涉及四个主要阶段：数据集增强、个性化反事实模型学习、反事实生成和框架评估。

A. 数据集增强

我们分析了 20 名大学生的数据，这些学生在 2020 年 COVID-19 封锁期间作为远程健康监测研究的一部分被监控了平均 7.8 个月。收集的数据包括来自 Oura 戒指和三星智能手表的睡眠指标、活动量、RMSSD 以及准备度得分，同时还收集了人口统计学数据和个人报告的情绪/心情数据。[14]

预处理涉及中位数插补、标准化（数值型）、独热编码（类别型）和频率汇总（情感/主题注释）。患者相似度是通过特征类型上的余弦相似性的加权平均值来计算的。然后，每个患者的时序数据被扩充了来自其三个最相似同伴的数据。患者的自有数据收到了更高的固定权重以强调个体特性，这一过程旨在稳定个性化模型，特别是在个人历史数据有限的情况下。此阶段的输出是为每位患者扩增后的数据集，包含 14 个生理特征。

B. 个性化反事实模型学习

为每位患者使用其增强数据集构建了一个个性化的反事实模型（PCM），将生理变量的变化 ($\Delta V_t = V_t - V_{t-1}$) 建模为滞后变量 (V_{t-1}) 的函数。

因果结构发现：我们使用了 PC 约束算法的时间适应版本 [15] 来识别从 V_{t-1} 到 ΔV_t 的合理预测关系。这通过使用加权偏相关，同时考虑来自增强状态的权重，迭代地检验滞后变量 ($V_{t-1,i}$) 和目标变化 ($\Delta V_{t,j}$) 之间的条件独立性。如果因果结构图中潜在影响的边的 p 值低于 0.05，则保留这些边。这为每个患者的每个 $\Delta V_{t,j}$ 产生了稀疏预测图。

模型训练：对于每个 ($\Delta V_{t,j}$)，训练了一个梯度提升机 (GBM) 来量化发现的关系。图 2 提供了一个来自单个患者的说明性示例，显示了这些学习到的关系以及所选预测因子对生理变化的影响程度。每个 GBM 的预测因子包括通过时间 PC 算法为该特定 $\Delta V_{t,j}$ 识别的滞后变量集 ($V_{t-1,i}$)，以及一个强制性的自滞后。在训练之前，预测变量和目标变量都使用来自患者增强训练数据的统计数据进行了 z 分数标准化。通过网格搜索和交叉验证优化 GBM 超参数。存储了来自每个 GBM 的标准化残差的标准差和标准化统计数据。

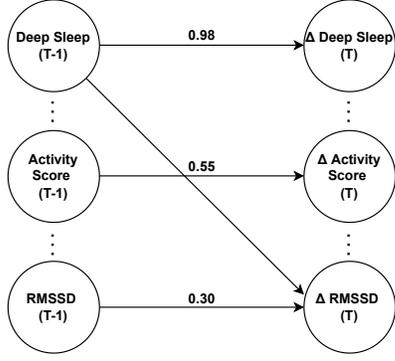


图 2: 来自单一患者的个性化模型的说明性预测结构。箭头描绘了滞后变量 V_{t-1} 对生理变化 ΔV_t 的学习影响。数值代表关键预测链接的相对特征重要性（源自 GBM）。

C. 反事实生成

学习的模型使得能够生成对假设干预的潜在结果。每位患者的验证后的 PCM（包括其训练过的 GBMs 和相关参数（标准化统计信息和学习到的噪声特征））推动了一个迭代过程。为了进行反事实生成，定义了初始生理状态 (X_0)，通常来自最近观测的数据。假设性干预（例如，修改“活动分数”）被应用于 X_0 。

生成过程在定义的时间范围内逐步展开。每一步 t :

- 1) 当前状态 X_t 已经标准化。
- 2) 预测了基于当前标准化状态的标准变化 (ΔZ_{pred})。
- 3) 从学习到的噪声分布中采样得到的一个随机元素 (ϵ_{t+1}) 被添加了。
- 4) 总的变化将标准化状态更新为 Z_{t+1} 。
- 5) Z_{t+1} 被非标准化为 X_{t+1}^{raw} ，被裁剪到现实的生理界限内，而 X_{t+1} 成为了下一次迭代的输入。

为了考虑噪声项带来的随机性，进行了蒙特卡罗模拟（多次运行使用不同的随机噪声样本），为每个场景生成了潜在未来轨迹的分布（反事实）。

D. 框架评估

整个框架通过评估模型准确性以及反事实的合理性和影响的指标来进行评估。最初，为了确保基础 PCM 准确捕捉短期动态，它们的一步预测准确性在保留数据上进行了验证。这包括生成确定性 ($\epsilon = 0$) 预测，并使

用平均绝对误差 (MAE) 和均方根误差 (RMSE)，计算预测值与实际值之间的差来量化准确性。随后，对反事实本身进行了评估。这涉及评估它们的可信性，定义为生成的时间点中 5th-95th 百分位范围内的轨迹（从蒙特卡罗运行）保持在预定义的实际生理边界内的比例。此外，还测量了基线偏移；对于给定的干预和结果，这量化了在地平线上生成的结果平均值（平均于蒙特卡罗运行）与基线值之间的差异。

III. 结果与讨论

我们评估了框架的一步预测准确性及其生成的反事实结果的可能性和影响。展示的结果来自使用 GBMs 训练的模型，这些模型利用了时间 PC 算法发现的关系，而没有明确的领域知识。

A. 模型性能和反事实可信度

个性化模型显示了捕捉短期生理动态的合理能力，具体预测性能指标详见表 I。例如，平均心率的平均绝对误差为 4.71 次/分钟，睡眠评分的平均绝对误差为 6.88 分。

表 I: 一步预测性能（均值（标准差））

Variable	MAE	RMSE
Average Heart Rate	4.71 (2.60)	5.67 (3.36)
Readiness Score (0-100)	6.45 (2.46)	7.69 (2.95)
Sleep Score (0-100)	6.88 (3.01)	8.02 (3.24)
Activity Score (0-100)	7.07 (3.26)	8.24 (3.76)

生成的生理轨迹表现出高度的反事实合理性（中位数 0.9643，四分位距：0.9643 – 1.0000 跨场景），轨迹大多保持在现实生理范围之内。

B. 干预的个性化效果

该框架的关键优势在于其能够捕捉到干预措施的不同反应。图 3 视觉上强调了对于假设的“高活动”和“低睡眠”场景，患者间的变异性。蜂群图中的每个点代表个别患者的预测关键结果变量 1 天的变化。

例如，在“高活动量”干预下，虽然平均“活动得分”显著增加（均值变化：+14.8 分），但个体反应差异很大。一些患者表现出轻微的变化，而另一些则显示出大幅增长。类似地，“低睡眠”干预通常会降低“睡眠得分”（均值变化：-16.0 分），但是这种减少的程度在不同个体间存在显著差异。这种异质性凸显了需要个性化

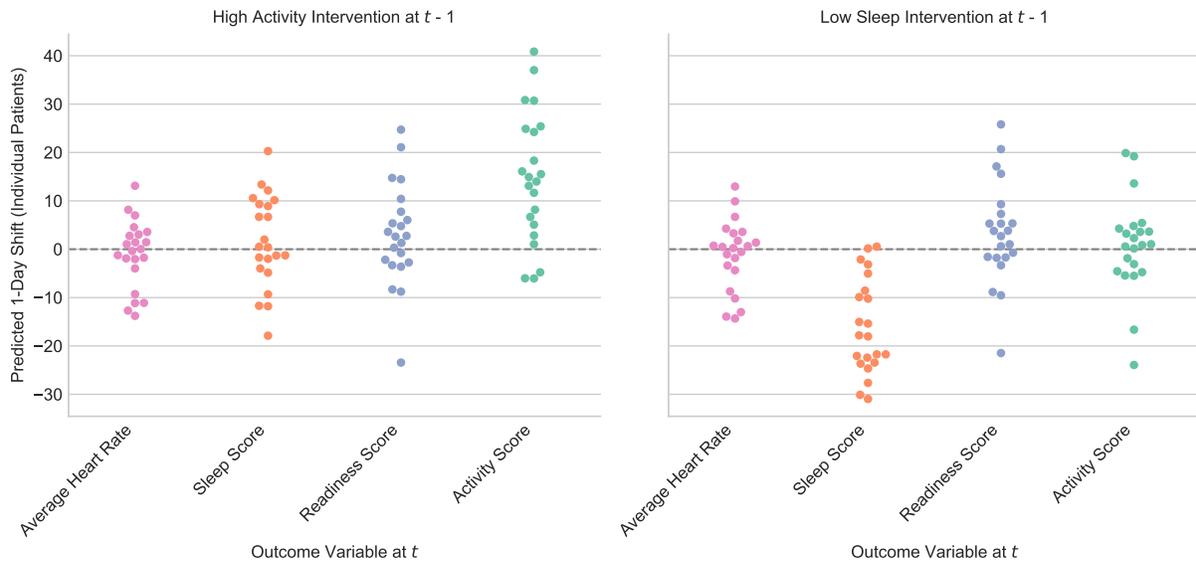


图 3: 异质性患者对干预措施的反应。每个点代表在“高活动”（左面板）和“低睡眠”（右面板）情景下，某一独特患者的预测 1 天变化量。群集图展示了干预效果在平均心率、睡眠得分、准备得分和活动得分上的个性化特征及变异性。虚线水平线表示零变化。

模型来预测个人对生活方式改变的反应，这对于制定有效的健康建议至关重要。总体而言，预测准确性、高度合理性以及个性化干预效果展示了我们的框架从可穿戴设备数据中生成可行见解的巨大潜力。

IV. 结论

我们提出了一种框架，利用可穿戴传感器数据创建个性化的反事实模型，以探索个体对假设干预的反应。我们的方法结合了患者相似性、日常生理变化的时间因果关系发现以及使用 GBMs（梯度增强机）进行预测建模，生成可能的结果，并突显干预效果中的显著个体间差异。我们通过准确率、合理性和与基线数据的偏差来评估这种方法，强调其超越通用见解向定制化健康策略转变的潜力。我们的工作的一个局限性是假设生理变量在时间上具有平稳性。未来的工作将专注于验证这些个性化的反事实模型对现实世界干预的有效性，扩展时间范围，并探索更复杂的动态交互以进一步提升可穿戴技术在个性化健康预测和决策支持方面的效果。

我们的框架具有更广泛的应用范围，超越了研究领域。例如，它可以作为轻量级云微服务运行在消费类可穿戴设备后面，将原始分数转化为个性化的“假设”提示，以鼓励健康习惯。此外，嵌入远程患者监测仪表盘

中时，它可以帮助临床医生预览干预效果。最终，通过将复杂的传感器数据流转化为具体行动的、针对特定患者的“假设”见解，这一框架为新一代主动式、真正个性化的数字健康干预奠定了基础。

参考文献

- [1] A. Kristofferson and M. Lindén, “A systematic review on the use of wearable body sensors for health monitoring: A qualitative synthesis,” *Sensors*, vol. 20, no. 5, p. 1502, 2020.
- [2] M. Haghi, S. Danyali, S. Ayasseh, J. Wang, R. Aazami, and T. M. Deserno, “Wearable devices in health monitoring from the environmental towards multiple domains: A survey,” *Sensors*, vol. 21, no. 6, p. 2130, 2021.
- [3] L. Hood and M. Flores, “A personal view on systems medicine and the emergence of proactive p4 medicine: predictive, preventive, personalized and participatory,” *New biotechnology*, vol. 29, no. 6, pp. 613–624, 2012.
- [4] A. Darwish and A. E. Hassanien, “Wearable and implantable wireless sensor network solutions for healthcare monitoring,” *Sensors*, vol. 11, no. 6, pp. 5561–5595, 2011.
- [5] M. Babu, Z. Lautman, X. Lin, M. H. Sobota, and M. P. Snyder, “Wearable devices: implications for precision medicine and the future of health care,” *Annual Review of Medicine*, vol. 75, no. 1, pp. 401–415, 2024.
- [6] P. Lago, “A state-of-the-art review of computational models for analyzing longitudinal wearable sensor data in healthcare,” *arXiv preprint arXiv:2407.21665*, 2024.

- [7] Z. Yang, I. Azimi, S. Jafarlou, S. Labbaf, J. Borelli, N. Dutt, and A. M. Rahmani, "Loneliness forecasting using multi-modal wearable and mobile sensing in everyday settings," in *2023 IEEE 19th International Conference on Body Sensor Networks (BSN)*. IEEE, 2023, pp. 1–4.
- [8] Z. Yang, Y. Wang, K. S. Yamashita, E. Khatibi, I. Azimi *et al.*, "Integrating wearable sensor data and self-reported diaries for personalized affect forecasting," *Smart Health*, vol. 32, p. 100464, 2024.
- [9] S. Jafarlou, J. Lai, I. Azimi, Z. Mousavi, S. Labbaf *et al.*, "Objective prediction of next-day's affect using multimodal physiological and behavioral data: Algorithm development and validation study," *JMIR Formative Research*, vol. 7, no. 1, p. e39425, 2023.
- [10] A. Subramanian, Z. Yang, I. Azimi, and A. M. Rahmani, "Graph-augmented llms for personalized health insights: A case study in sleep analysis," in *2024 IEEE 20th International Conference on Body Sensor Networks (BSN)*. IEEE, 2024, pp. 1–4.
- [11] Y. Raita, C. A. Camargo Jr, L. Liang, and K. Hasegawa, "Big data, data science, and causal inference: a primer for clinicians," *Frontiers in Medicine*, vol. 8, p. 678047, 2021.
- [12] P. Sanchez, J. P. Voisey, T. Xia, H. I. Watson, A. Q. O' Neil, and S. A. Tsafaris, "Causal machine learning for healthcare and precision medicine," *Royal Society Open Science*, vol. 9, no. 8, p. 220638, 2022.
- [13] R. Moraffah, P. Sheth, M. Karami, A. Bhattacharya, Q. Wang, A. Tahir, A. Raglin, and H. Liu, "Causal inference for time series analysis: Problems, methods and evaluation," *Knowledge and Information Systems*, vol. 63, pp. 3041–3085, 2021.
- [14] S. Labbaf, M. Abbasian, B. Nguyen, M. Lucero, M. S. Ahmed, A. Yunusova, A. Rivera, R. Jain, J. L. Borelli, N. Dutt *et al.*, "Physiological and emotional assessment of college students using wearable and mobile devices during the 2020 covid-19 lockdown: An intensive, longitudinal dataset," *Data in Brief*, vol. 54, p. 110228, 2024.
- [15] M. Kalisch and P. Bühlman, "Estimating high-dimensional directed acyclic graphs with the pc-algorithm." *Journal of Machine Learning Research*, vol. 8, no. 3, 2007.