

分类号
U D C

密级
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 基于加权核范数矩阵填充的因果
效应评价方法研究

研究生姓名: 邢晓文

指导教师姓名、职称: 牛成英、教授

学科、专业名称: 统计学、数理统计学

研究方向: 复杂数据分析

提交日期: 2024年6月5日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 邢晓文 签字日期： 2024年6月3日

导师签名： 毕成英 签字日期： 2024年6月3日

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意 (选择“同意”/“不同意”)以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊(光盘版)电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 邢晓文 签字日期： 2024年6月3日

导师签名： 毕成英 签字日期： 2024年6月3日

Research on Causal Effect Evaluation Based on Weighted Kernel Norm Matrix Completion

Candidate : Xing Xiaowen

Supervisor: Niu Chengying

摘 要

随着大数据科学发展以及因果推断相关研究的不断推进,从观察数据中发现因果关系是各领域研究中的热点问题。在反事实因果效应估计框架下,潜在结果的估计是关键问题。目前,随着机器学习功能变得越来越强大,越来越多的机器学习算法融入因果推断方法中,利用矩阵填充方法填补缺失潜在结果已经被证实是因果推断中一种有效方法。从矩阵填充角度出发,研究在因果推断中如何更好的利用先验信息降低矩阵填充估计潜在结果的误差。主要研究内容包括以下几方面:

(1)在不考虑协变量的情况下,填充目标变量的反事实结果。提出用自适应加权核范数替代核范数的加权核范数矩阵填充方法,避免了较大奇异值过度惩罚的情况,并根据交替方向乘子法(Alternating Direction Method of Multipliers, ADMM)框架证明最优解以及收敛性。从数值模拟和真实数据实验结果来看,相较于传统方法加权核范数模型预测缺失潜在结果精度更高。

(2)在考虑协变量的情况下,提出基于随机森林倾向得分加权矩阵填充的方法,用随机森林模型估计倾向得分能有效避免传统倾向得分方法模型设定匹配结果的平衡性,在此基础上,利用倾向得分匹配协变量平衡性质加权损失函数,使得处理组和控制组之间具有可比性。从实验结果来看,随机森林倾向得分加权矩阵填充模型估计潜在结果的误差更小,并估计出因果推断关注的平均处理效应。

关键词: 因果推断 矩阵填充 加权核范数 随机森林 倾向得分

Abstract

With the development of big data science and the continuous advancement of research on causal inference, discovering causal relationships from observational data is a hot topic in various fields of research. Under the framework of counterfactual causal effect estimation, the estimation of potential outcomes is a key issue. Currently, as machine learning capabilities become increasingly powerful, more and more machine learning algorithms are being integrated into causal inference methods. The use of matrix completion methods to complete in missing potential outcomes has been proven to be an effective method in causal inference. This article starts from the perspective of matrix completion and studies how to better utilize prior information in causal inference in causal inference to reduce the error of potential outcomes in matrix completion estimation. The main research content includes the following aspects:

(1) Without considering covariates, complete the counterfactual results of the target variable. This article proposes a weighted kernel norm matrix completion method, which replaces the kernel norm with an adaptive weighted kernel norm to avoid excessive punishment for large singular values. The optimal solution and convergence are proven using the ADMM framework. From the results of numerical simulation and real data experiments, it can be seen that compared to traditional methods, weighted kernel norm models have higher accuracy in predicting potential

missing outcomes.

(2) Taking into account covariates, this article proposes a method based on random forest propensity score weighted matrix completion model. Estimating propensity scores using a random forest model can effectively avoid the balance of matching outcomes model settings in traditional propensity score methods. Based on this, weighted loss function using propensity score covariate balance property to ensure comparability between treatment group and control group. From the experimental results, it can be seen that the random forest propensity score weighted matrix completion model has a smaller error in estimating potential outcomes and estimates the average treatment effect for the treated in causal inference.

Keywords: Causal inference; Matrix completion; Weighted kernel norm; Random forest; Propensity score

目 录

1 绪论	1
1.1 研究背景与研究意义.....	1
1.1.1 研究背景.....	1
1.1.2 研究意义.....	1
1.2 国内外研究现状.....	2
1.2.1 国外研究现状.....	2
1.2.2 国内研究现状.....	5
1.3 主要研究内容和重点解决的问题.....	8
1.3.1 主要研究内容.....	8
1.3.2 重点解决的问题.....	8
1.4 论文组织结构.....	9
1.5 创新点.....	10
2 反事实框架下的因果推断方法	11
2.1 反事实框架.....	11
2.1.1 反事实结果.....	11
2.1.2 重要假设.....	12
2.1.3 因果效应.....	13
2.2 经典因果推断方法.....	13
2.2.1 双重差分法.....	14
2.2.2 合成控制法.....	14
2.2.3 因子估计方法.....	15
2.3 基于机器学习的因果推断方法.....	15
2.3.1 弹性网络.....	15
2.3.2 核范数矩阵填充方法.....	16
2.4 小结.....	16
3 基于加权核范数最小化的矩阵填充反事实结果估计	18

3.1 引言.....	18
3.2 理论基础.....	19
3.3 加权核范数矩阵填充反事实结果估计.....	20
3.3.1 加权核范数最小化模型.....	20
3.3.2 最优解证明.....	21
3.3.3 模型优化.....	24
3.4 数值检验.....	27
3.4.1 模拟数据生成.....	27
3.4.2 方法模拟与精度比较.....	27
3.5 反事实结果估计.....	31
3.5.1 数据来源.....	31
3.5.2 反事实结果分析.....	31
3.6 小结.....	32
4 基于倾向得分加权矩阵填充的因果效应分析.....	33
4.1 引言.....	33
4.2 方法介绍.....	34
4.2.1 倾向得分.....	34
4.2.2 随机森林倾向得分模型.....	35
4.2.3 倾向得分加权损失函数矩阵填充模型.....	36
4.3 数值模拟.....	37
4.3.1 数据生成.....	37
4.3.2 随机森林估计倾向得分.....	37
4.3.3 因果效应分析.....	39
4.4 基于倾向得分加权的因果效应分析.....	40
4.4.1 数据来源与处理.....	40
4.4.2 倾向得分平衡协变量.....	40
4.4.3 平均处理效应估计.....	42
4.5 小结.....	42
5 总结与展望	44

5.1 总结.....	44
5.2 展望.....	44
参考文献.....	46
攻读硕士学位期间承担的科研任务及主要成果.....	52
致谢.....	53

1 绪论

1.1 研究背景与研究意义

1.1.1 研究背景

在过去的 30 多年中，评估因果效应已成为统计学和计量经济学领域的重要研究方向。因果关系识别一般分为观察和实验两种方法。目前，随机对照试验方法被视为最有效的实验性研究方法，认为是确定因果效应的黄金标准，其假设在试验中不存在混杂因素，关联即因果关系。然而，在实践中进行随机试验往往面临着诸多困难，甚至是不切实际的。因此，在这种情况下，研究者必须依赖于观测数据，通过分析这些观测数据来揭示因果关系。

对于观测性数据，核心问题是如何得到反事实结果。在观察性研究中，无论是对接受处理的个体还是对不接受处理的个体，都无法同时观测到单个个体两种潜在结果，导致数据缺失。因此，因果效应难以推断在于始终存在一个潜在结果无法观测到的问题。然而，利用机器学习方法通过求解优化问题对缺失数据矩阵进行填充，以此来估计潜在结果，进而实现因果效应的估计，成为了各界学者研究和关注的重点。

1.1.2 研究意义

在因果推断应用中缺失来自于处理的选择，并且缺失数据有特定结构，例如块状、条状缺失、交错缺失。传统处理随机缺失机制的数据插补方法已不能很好地处理政策评估过程中数据缺失问题。矩阵填充问题一般在给定一个部分观察到的矩阵，估计缺失元素。矩阵填充通常考虑随机缺失数据的模式，矩阵填充技术在推荐系统、视频去噪和图像恢复等领域有着广泛的应用。特别是对条块缺失数据修复有其独特优势。因此，研究矩阵填充在因果效应中的应用具有重要意义。

由于混杂因素会引起因果识别结果存在偏差，因此处理混杂因素是因果推断中的一个重要挑战。为了解决这一挑战，倾向得分法应运而生，其旨在消除混杂因素对观察数据平均处理效应的影响。通过正确的倾向得分可以平衡变量分布，

从而揭示观测数据中的真实因果关系。利用机器学习方法代替传统方法估计倾向得分，可以有效地避免倾向得分模型设定错误的情况。因此，研究随机森林估计倾向得分为观察数据的因果效应估计提供新思路。

1.2 国内外研究现状

1.2.1 国外研究现状

(1)因果推断

因果推断的关键在于构建反事实结果，即与观测到的情况相反的结果。由于研究者无法同时观测处理和控制状态下的研究对象，为了估计处理对结果的效应，研究者通常利用随机分配处理来创建一个与处理组尽可能相似的控制组，通过比较在结果变量上的平均差异，就可以揭示因果效应。

随着因果推断相关研究的推进，现有估计因果效应的方法有双重差分法(Difference in Difference, DID)、合成控制法(Synthetic Control Methods, SCM)、基于因子估计方法和机器学习方法等。

DID 是用于估计因果效应的常用方法之一，在实证中有非常广泛的应用。DID 是用双向固定效应(Two Way Fixed Effect, TWFE)来估计平均处理效应，Chaisemartin 等(2020)指出使用 TWFE 估计得到各组、各时期的平均处理效应的加权总和。当处理效应存在异质性时，负权重会导致最终计算的总平均处理效应呈现负值。Callaway 等(2021)考虑评估因果效应的时变特征，观测到数据个体为多期，并且处理对个体的影响可能在不同时间存在交错的情况。此外，假设只有在对一些可观测的解释变量进行调整之后，才能建立起有效的因果关系。

Abadie 等(2003)首次提出的 SCM 侧重于数据矩阵仅存在单行缺失值的情况。SCM 就使用处理组作为数据基础，“合成”出一个在其他方面类似于处理组样本的反事实样本，来考察处理结果。另外，SCM 综合考虑个体和时间两个维度上存在共同因素交互的情况，即个体 i 在未受处理时的潜在结果是由可观测的协变量和不可观测的共同因子共同决定。在 Abadie 等(2010)研究基础上，Amjad 等(2017)采用奇异值阈值方法去除原始数据中的噪声，放宽了对凸组合限制，并采用更一般的线性组合来合成控制组的反事实结果。同样，Ben-Michael 等(2021)也

松弛了对凸组合的约束，并在权重估计中使用岭回归以提高拟合效果。然而，Abadie(2021)指出，放宽对凸组合的假设可能导致外推偏误，因此，Abadie 和 L'Hour(2021)在目标函数中引入了惩罚项，以确保控制组个体和受处理个体特征匹配程度，以此降低插值偏误。Ben-Michael 等(2021)将 SCM 推广到更广泛的应用场景，考虑了多个体和多时期受到处理的情况。

虽然 SCM 具有因子结构，但它并未直接估计因子模型，而是通过对控制组因变量进行加权组合来估计处理组的反事实状态。在基于因子估计方法中，构建反事实时首先需要估计出因子模型中的因子个数、因子和载荷，然后再估计因变量的预测值来构建反事实。在大多数研究中采用渐进主成分分析法来估计因子模型，Pesaran(2006)采用两步回归法得到的估计量不是一致的，Bai 等(2009)通过迭代的方式得到收敛的一致估计量。为解决两个潜在因变量不可同时被观察的问题，Bai 等(2021)从面板处理效应的实质出发，采用因子模型进行矩阵填充，以构建反事实结果。这种方法灵活处理当处理效应存在异质性的情况，也能考虑处理组只有一个个体的情况。Bai 等(2021)的研究关注矩阵块状缺失的情况，然而在实际应用中常见随机缺失现象。Jin 等(2021)则研究因子模型存在随机缺失的情况下的应用，并使用期望最大化(Expectation Maximum, EM)算法来获得因子和载荷的一致估计。在已有的机器学习方法中，Athey 等(2021)提出通过求解正则化核范数的优化问题估计反事实，而不是一致的估计因子和载荷，是将矩阵填充方法应用于估计因果效应。

如前所述，每个个体只会处于被处理影响或不受处理影响的状态，这使得因果效应难以推断的核心问题在于某种状态的缺失问题。因此，利用机器学习的方法通过求解最优化问题来填充数据矩阵中的缺失值，即缺失的潜在结果，从而实现因果效应的估计和推断。

(2)矩阵填充

矩阵填充问题近年来成为研究热点之一，旨在根据少量观测值精确地重构原始矩阵。整个问题可以看作是一个带有结构性假设的优化问题。在常见的矩阵完填充方法中，通常采用低秩结构假设，即假设高维矩阵的行或列是由少量行或列隐含生成的。在实际应用中，著名的 Netflix 推荐系统就是矩阵填充的一个典型例子。同样在图像恢复、视频去噪等领域矩阵填充都发挥着重要作用。

Candès 和 Recht(2009)提出用核范数作为罚函数使得整个优化问题变得更容易计算,不再是 NP-难。并假设完整矩阵是低秩矩阵加上噪声,以及缺失是完全随机的。目前为解决核范数最小化问题,已经拥有大量的快速算法。如 Fazel 等(2003)提出的半正定规划算法,通过半正定规划程序求解优化问题,Toh 等(2010)提出了加速近似梯度法,与此同时,还有奇异值阈值算法(Cai 等, 2010)、加速奇异值阈值法(Hu 等, 2012)以及不经过奇异值分解的快速奇异值阈值法(Cai 等, 2013)等。

当数据是随机缺失机制时,结果变量是否被观测到的概率取决于一些可观测的协变量,而与其具体取值无关。然而,在绝大多数情况下,无法获取观测概率矩阵的先验知识,因此需要先估计观测概率矩阵,再将其代入损失函数中。Mao 等(2019)假设观测概率取决于协变量矩阵,即符合随机缺失机制,利用 Frobenius 范数和核范数正则化提出估计的渐进收敛性。同时,Wang 等(2021)使用非参数模型,寻找合适的权重矩阵代替缺失概率矩阵,得到的权重既不依赖缺失机制,而且不依赖观测矩阵。

当数据是非随机缺失时,即数据缺失与否取决于结果变量的具体取值,这有违随机缺失机制。Schnabel 等(2016)对一维观测值采用逆概率加权,进行矩阵填充。然而,在实际应用中经常出现较大方差。因此,Wang 等(2019)设计了双稳健估计对缺失矩阵进行加权。

如上述文献中的矩阵填充方法可以解决多样化的缺失数据问题,适用于不同的数据特征和应用场景。一些方法可能存在扩展性问题或很大程度上取决于关于数据分布的基本假设,而这些假设在实践中可能并不总是成立的。虽然每种方法都有其优势和局限性,但从观察性数据出发,进一步研究矩阵填充方法具有现实意义。

(3)倾向得分加权

平均处理效应是因果推断中评估某个处理变量对结果变量产生的影响,匹配法、逆概率加权法、双稳健法等都是通过一个共同的中间量——倾向得分来实现反事实结果的估计。倾向得分首次由 Rosenbaum 和 Rubin(1983)提出,指在一定的协变量条件下,观察个体接受干预的概率。因此,倾向得分可以最大限度地概括协变量对缺失的影响。在计算出倾向得分后,可采用加权法来纠正缺失的影响。

在许多情况下,因为样本接受处理通常是非随机选择的,所以再进行严格的随机对照试验往往面临一定困难。此外,处理组和控制组的协变量分布不一致可能导致因果效应的估计出现较大偏差。许多研究采用协变量平衡法来修正组间协变量分布的差异,以减少估计误差。其中一种方法是参数方法,如 Imai 和 Ratkovic(2014)提出通过逆概率处理加权法(Inverse Probability of Treatment Weighting, IPTW),得到处理组和控制组的协变量在加权后的一阶矩相同的倾向得分参数估计量。由于 IPTW 估计量对倾向得分模型的错误分类敏感, Fan 等(2022)通过直接优化平衡预处理协变量来估计倾向得分。Shen 等(2018)提出因果正则逻辑回归算法,基于样本学习权重得到选择性偏差更稳健的预测模型。Li 等(2018, 2019)将方法拓展至多元处理变量的情况。Cheng 等(2019)基于协变量最小损失提出使用 Lasso 估计倾向得分。Ning 等(2020)研究在潜在混杂因素数量多于样本量时,提出了一种稳健的高维协变量平衡倾向得分法估计平均处理效应。另一种方法是非参数方法,由于倾向得分仅出协变量平衡的中间环节出现,Josey 等(2021)跳过估计倾向得分的步骤,而是通过求解约束优化问题来估计平衡权重以此直接估计权重。

倾向得分加权方法能在观测数据中有效地控制混杂因素的影响,适用于处理复杂数据和潜在因素较多的情况。但倾向得分加权方法对倾向得分模型设定和样本量要求较高,并且对匹配结果的平衡性要求严格。因此,利用机器学习的方法避免倾向得分模型设定问题,提高匹配的平衡性具有重要研究价值。

1.2.2 国内研究现状

(1)因果推断

因果推断作为数据科学和社会科学领域的重要分支,在学术界受到广泛关注。董珍(2019)将双重政策因果效应评估方法推广到更一般的多政策情形,提出了多重政策下各政策因果效应识别条件以及相应的估计方法。韩锋(2020)给出在协变量缺失时倾向得分的估计方法,并应用倾向得分扩展加权给出因果效应的估计。丁杨(2021)梳理和总结了因果推断的三种常见的方法:格兰杰因果推断、倾向得分匹配和双重差分法。机器学习在因果推断领域的应用具有重要的实际意义。钱浩祺等(2021)基于机器学习视角文从因果推断的两大主流分析框架出发,从匹配

法、双重差分法和合成控制法等方法出发,详细地阐述了机器学习在提升因果效应识别方面的理论基础。李超等(2021)提出机器学习和传统计量方法结合以及面板数据的因果识别是未来研究的热点。欧祥威等(2022)通过因果网络描述大数据背景下因果机制的可能性。而蒋青嬗等(2023)则利用倾向得分构建处理组和控制组协变量的经验加权分布,通过最小化分布差异来最大化协变量平衡,从而估计倾向得分和平均处理效应。

国内学者关于因果推断的研究与国外主流研究基本一致。提出的基于机器学习的因果推断方法能更好地处理大规模数据和非线性关系。在大数据背景下,从矩阵填充角度出发估计因果推断所关注的潜在结果,是众多学者关注的热点问题具有重要研究意义。

(2)矩阵填充

国内关于矩阵填充的研究在多方面取得显著成就。协同过滤是推荐系统最常用的方法之一,矩阵填充在协同过滤中扮演着重要的角色,帮助改善推荐系统的性能。为解决数据稀疏性问题,张玮奇等(2015)提出加权核范数最小化模型。王志远等(2020)考虑选择倾向度改进填充矩阵的 Slope One 算法。孙乐等(2020)提出矩阵填充与 BP 神经网络的混合推荐算法。张晨等(2021)提出混合协同过滤推荐算法,利用随机梯度下降方法对结果进行分析,以寻找最佳参数。潘伟等(2021)从贝叶斯框架出发,提出解决低秩矩阵问题的分层高斯先验模型。这些算法都能够有效地缓解数据稀疏性问题,并且相较于传统推荐算法,它们在提高推荐准确度方面表现出显著的优势。

为了解决在矩阵填充过程中高维度和高计算成本问题,冯雅莉等(2019)提出基于快速随机投影的矩阵填充方法,该方法利用随即投影对填充矩阵进行降维,然后使用奇异值分解的近似模型重构矩阵,以实现矩阵填充。基于增广拉格朗日乘子法,王俊霞等(2021)提出的修正算法对每步产生的阈值进行符号投影,形成新符号矩阵,同时证明在合理条件下,当罚因子充分大时,修正算法是收敛的。肖云等(2022)基于均值的增广拉格朗日乘子(MALM)算法,提出了一种尾端修正的 Toeplitz 矩阵填充新算法,在一定程度上减少了每步均值处理所产生的数据传输量,从而降低了计算代价。温瑞萍等(2022)提出快速且具有较高精度的 Toeplitz 矩阵填充算法,通过在迭代过程中嵌入修正步而极大地节约了计算时间,得到了

更精确的填充矩阵。

张娅楠等(2016)提出基于 Schatten-p 范数最小化的弹性网正则化模型解决复杂问题解的稳定性。当数据矩阵变大且观测数据较少时,容易导致过拟合,邝神芬等(2019)提出的基于深度矩阵分解网络的矩阵填充方法不仅能弥补传统矩阵分解的缺点,而且能处理复杂的非线性数据。石莹等(2022)利用加权核范数去松弛原始低秩极小化问题,基于 Soft-Impute 算法思想提出 WNNM-Impute 算法,有比 Soft-Impute 和大多数对比算法更快的收敛速率。薛娇等(2022)采用函数型数据分析方法,构造基于多视角学习的非负函数型矩阵填充算法。潘伟等(2023)提出了一种基于加权 Schatten-p 范数最小化模型对矩阵进行低秩约束。

综上所述,国内的大多数研究是在国外的基础上进行方法改进,提出了用于不同场景和数据类型的新模型和算法。然而,随着机器学习算法的不断融合,利用矩阵填充方法填补缺失潜在结果是研究因果推断的有效方法之一。

(3)倾向得分加权

倾向得分加权方法在不同学科和应用领域受到广泛关注。刘展等(2016)针对非概率抽样统计推断问题,提出采用倾向得分匹配选择样本,再对匹配样本进行加权调整来估计目标总体。邵凌楠等(2018)在随机缺失机制假定下,提出逆概率多重加权估计方法。彭非等(2019)提出了是将基于欧氏距离匹配法的加权自助法推广至倾向性得分匹配法,解决重抽样样本中个体被匹配数不是原样本的一致估计问题方法。吴浩等(2020)在 Li 等(2018)提出的协变量加权平衡法基础上,提出了基于协变量平衡加权法的稳健且有效的估计方法,并通过引入超级学习算法提升了模型在实证应用中的稳健性。高倩等(2021)从变量选择和协变量平衡两个角度出发,总结高维情况下基于倾向得分的因果推断方法。

针对国内特有的数据结构和特点,基于大数据和机器学习的倾向得分加权方法应用领域广泛多样。然而,在使用倾向得分加权方法时需要处理样本选择偏差、数据不平衡等问题。因此,利用机器学习方法克服以上问题,提供更可靠和稳健的结果。

1.3 主要研究内容和重点解决的问题

1.3.1 主要研究内容

本文主要研究内容包括两部分，分别是：

(1) 基于加权核范数最小化的矩阵填充反事实结果估计

该部分主要从矩阵填充角度出发估计因果推断中缺失的潜在结果。已有的矩阵填充算法估计潜在结果依赖于奇异值分解，针对所有奇异值使用同一值收缩问题，提出加权核范数最小化的矩阵填充模型考虑对较大的奇异值进行较小惩罚，而对较小奇异值进行较大惩罚。此外还需通过数值检验和实证检验分析验证方法可行性。数值检验部分通过随机生成相互独立的高斯随机变量，对不同规模和缺失率的数据矩阵进行矩阵填充；实证分析部分将加权核范数矩阵填充方法应用于加州控烟真实数据，与传统因果效应估计方法就估计精度进行比较。

(2) 基于倾向得分加权矩阵填充的因果效应分析

该部分主要考虑充分利用协变量信息来估计因果效应，由于协变量间存在异质性，在因果推断中常用倾向得分来消除混杂因素的影响。倾向得分是在给定协变量的情况下样本进行处理组的概率。在因果推断中已提出的在矩阵填充模型对所有观测到的差异赋予相同的权重，而矩阵填充的目的是根据观测数据预测相似的缺失的潜在结果，因此利用倾向得分加权矩阵填充模型在理论上是可行的。重点在于估计倾向得分，为避免传统方法要求模型正确设定问题，使用随机森林算法估计倾向得分可有效避免这一问题。同样，通过数值检验和实证检验分析验证方法可行性。数值检验部分通过随机生成符合条件的协变量，利用随机森林算法估计倾向得分，最后将估计的倾向得分加权矩阵填充模型，以此估计反事实结果；实证分析部分在考虑协变量的情况下，将倾向得分加权矩阵填充方法应用于包含协变量地加州控烟数据，与传统因果效应估计方法进行比较。

1.3.2 重点解决的问题

从机器学习视角出发进行政策评估，利用矩阵填充方法以期估计潜在结果，从而计算因果效应。

(1)在不考虑协变量影响的情况下，加权核范数最小化的矩阵填充方法估计反事实结果，利用增广拉格朗日乘子算法求解目标函数，根据辅助函数法证明其收敛性。

(2)在考虑协变量影响的情况下，倾向得分加权损失函数的矩阵填充方法估计潜在结果，利用倾向得分来平衡样本间权重达到协变量平衡效果。

(3)将基于加权核范数最小化的矩阵填充方法和基于倾向得分加权损失函数的矩阵填充方法应用于真实数据集上。

综合以上研究内容，本节研究思路与技术路线如图 1.1 所示。

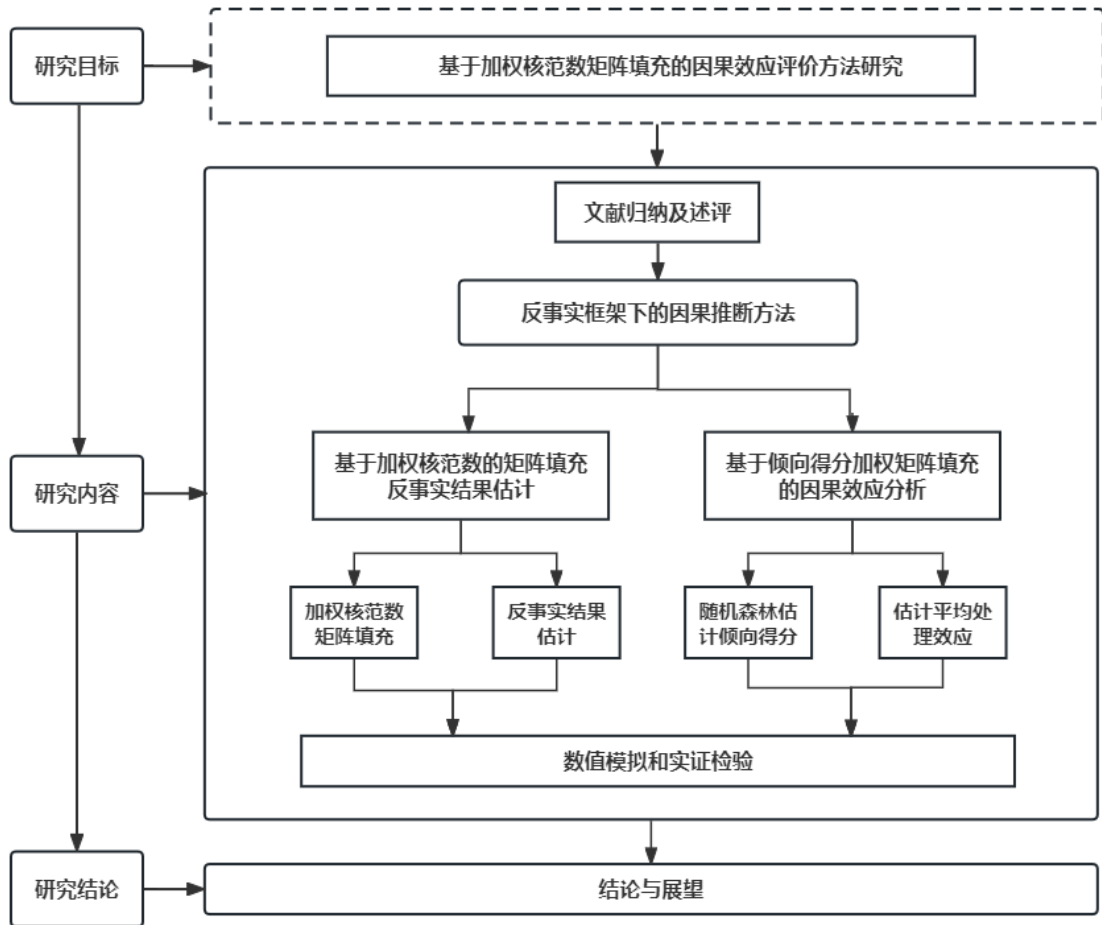


图 1.1 研究思路与技术路线

1.4 论文组织结构

从矩阵填充角度出发，研究因果推断的核心内容估计潜在结果。通过对矩阵

填充方法和倾向得分估计方法进行改进,能更有效估计因果效应。最后进行实证检验解决实际问题。主要安排如下:

第1章为绪论,主要概括了利用矩阵填充模型在因果推断中的研究背景及意义,并通过梳理文献介绍近几年国内外学者对因果推断、矩阵填充和加权倾向得分方面的研究现状,提出的研究内容和组织框架,以及创新点。

第2章主要介绍在反事实框架下的因果推断方法。首先介绍反事实框架,然后介绍双重差分法、合成控制法和因子估计方法的经典方法,最后介绍了弹性网络和核范数矩阵填充方法。

第3章在不考虑协变量的情况下,对现有的矩阵填充方法进行改进,用加权核范数代替核范数以保留更多有用信息,通过ADMM框架证明其最优解和收敛性,并进行了数值模拟和实证检验。

第4章将倾向得分的协变量平衡性质消除变量间的异质性,使观测数据具有可比性,并提出基于随机森林倾向得分加权矩阵填充模型,利用加州控烟数据进行实证分析,并估计出了平均处理效应。

第5章对主要研究内容进行了总结,并根据现有方法的不足提出后续研究的方向。

1.5 创新点

(1)因果效应难以推断的本质问题是潜在结果缺失的问题,将估计缺失元素的矩阵填充方法应用于因果推断中,估计受到干预个体的反事实结果,从而计算平均处理效应。

(2)在核范数矩阵填充的因果推断方法基础上,提出基于加权核范数最小化矩阵填充方法估计反事实结果,解决对所有奇异值用同一值收缩问题。在考虑协变量的情况下,利用随机森林估计倾向得分,能有效避免传统倾向得分模型误设,利用倾向得分匹配协变量平衡性质加权矩阵填充损失函数估计潜在结果,使得处理组和控制组之间具有可比性,并提高预测精度。

2 反事实框架下的因果推断方法

2.1 反事实框架

Neyman-Rubin 反事实框架认为，个体在处理和控制在状态中都有潜在结果，即在观测到的结果和未被观测到的结果。在因果推断中，反事实估计框架已经成为随机试验和观察性研究中推断因果效应的基准。

2.1.1 反事实结果

因果推断的核心是比较原因变量对结果的影响。假设在 T 时期内观测到 N 个个体的面板数据中，用虚拟变量 W_{it} 表示第 i 个个体在时间 t 的处理状态，其中 $i=1,2,\dots,N$ ， $t=1,2,\dots,T$ 。通常假设处理变量为二元处理变量，即对于任意个体 i 均存在两种处理状态： $W_{it}=1$ 表示受到处理（处理组）， $W_{it}=0$ 表示没有受到处理（控制组），每种处理状态对应一个潜在结果用 $Y_{it}(0)$ 和 $Y_{it}(1)$ 表示。反事实定义为在原因或处理不存在时，个体会发生的潜在结果。可以用表 2.1 表示反事实框架。

表 2.1 反事实框架

	$Y_{it}(1)$	$Y_{it}(0)$
控制组 ($W_{it}=0$)	反事实结果	观测结果
处理组 ($W_{it}=1$)	观测结果	反事实结果

由表 2.1 可知，对于控制组的个体来说，反事实就是当个体处于处理状态下的潜在结果。相对应地，对于处理组的个体而言，反事实就是当个体处于控制状态下的潜在结果。观测到的潜在结果称为观测结果，而未观测到的潜在结果称为反事实结果。在实际观测数据中无法观测到反事实结果，那么反事实结果本质上是一个缺失值。

因此，因果效应实际上也是缺失数据问题，解决该问题的关键在于利用已知的信息对无法观测到的潜在结果进行估计。表 2.1 可以用数学模型表示为：

$$Y_{it} = W_{it}Y_{it}(1) + (1 - W_{it})Y_{it}(0) \quad (2.1)$$

其中,观测到的结果 Y_{it} 取决于二元处理变量 W_{it} ,当 $W_{it} = 1$ 时, $Y_{it}(1)$ 为观测结果, $Y_{it}(0)$ 就是反事实结果;当 $W_{it} = 0$ 时, $Y_{it}(0)$ 为观测结果, $Y_{it}(1)$ 就是反事实结果。

2.1.2 重要假设

假设 2.1: 稳定性假设即稳定个体处理值假设(Stable Unit Treatment Value Assumption, SUTVA)包含两层含义:

(1)每个个体的潜在结果不依赖于其他个体的处理状态,即不同个体的潜在结果间不会有交互效应;

(2)处理水平对每个个体都相同。

在反事实框架中,稳定性假设第一个要求保证对于个体 i 处理状态只有两个且只有两个潜在结果;第二个要求每个个体受到的处理水平相同。然而,在实际应用中处理的同质性很难保证,因此更加关注稳定性假设的第一项要求。

分配机制是识别因果关系的关键。在随机化试验中,分配机制是人为控制的,是已知的。在观察性研究中,分配机制是未知的。然而,在观察性研究的目的是识别未知的分配机制,从而估计因果效应。为了更准确表达反事实结果估计,需要引入非混杂性的概念,也称为条件独立性或根据观测变量选择。非混杂性是指控制变量 X_{it} 后,个体处理状态分配独立于潜在结果,可以表示为:

$$(Y_{it}(0), Y_{it}(1)) \perp\!\!\!\perp W_{it} \mid X_{it}, \quad (2.2)$$

其中, $\perp\!\!\!\perp$ 表示相互独立, X_{it} 表示协变量。式 2.2 表示以 X_{it} 为条件, 处理变量 W_{it} 独立于潜在结果 $Y_{it}(0)$ 和 $Y_{it}(1)$ 以及潜在结果的任何函数。

假设 2.2: 强可忽略性假设(Strongly Ignorable Treatment Assignment), 若满足:

$$(Y_{it}(0), Y_{it}(1)) \perp\!\!\!\perp W_{it} \mid X_{it}, \quad 0 < P(W_{it} = 1 \mid X_{it}) < 1,$$

那么称处理分配机制是可忽略的。

满足可忽略性假设的第一个条件即非混杂性，处理分配与潜在结果间相互独立，因此也称非混杂性为条件独立性。第二个条件即共同区间假设，则要求根据协变量 X_{it} 分层后，接受处理的概率大于 0，保证每层内均有处理组和控制组的个体，从而保证可以估计因果效应参数。

2.1.3 因果效应

在实际问题的研究中，因果效应是反应同一变量在不同处理下潜在结果间的比较，往往关注处理的平均因果效应，而不是某特定个体的因果效应。

常用总体平均因果效应(average treatment effect, ATE)定义所有个体因果效应的平均值，表示从一个总体中随机抽取一个个体进行处理的平均因果效应：

$$\tau_{ATE} = E[Y_{it}(1) - Y_{it}(0)],$$

受处理影响的个体的平均因果效应称为处理组平均因果效应(average treatment effect of the treated, ATT)表示为：

$$\tau_{ATT} = E[Y_{it}(1) - Y_{it}(0) | W_{it} = 1],$$

没有受处理影响的个体如果结果处理的平均因果效应称为控制组平均因果效应(average treatment effect of the control, ATC)表示为：

$$\tau_{ATC} = E[Y_{it}(1) - Y_{it}(0) | W_{it} = 0].$$

因为两个所需的潜在结果总有一个是缺失的，所以在任何情况下因果效应都是不可识别的。反事实框架认为可以通过未受处理个体的平均结果 $E(Y_{it}(0) | W_{it} = 0)$ 来估计反事实 $E(Y_{it}(0) | W_{it} = 1)$ ，以此来估计因果效应。以下假设在观察性研究中，可实现对处理效应的一致估计。

2.2 经典因果推断方法

在反事实框架下，利用面板数据估计处理效应的常用方法主要包括：双重差分法(DID)、合成控制法(SCM)和因子估计方法。大多数研究中采用经典方法进行对比，评价所提出方法的效果和性能来增强可信度，以下介绍常用的经典因果推断方法。

2.2.1 双重差分法

双重差分法关注的是因变量在处理前后的变化,假设处理组和控制组具有相同的线性趋势,经过两次差分就可以得到平均处理效应。通常经典的双重差分法通过 TWFE 模型进行估计。因此,可以得到可观测的因变量模型为:

$$Y_{it} = \beta_0 + \beta_1(G_i \times T_t) + \beta_2^\top x_{it} + \delta_i + \gamma_t + \varepsilon_{it}, \quad (2.3)$$

其中, G_i 表示个体属于控制组 ($G_i = 0$) 还是处理组 ($G_i = 1$); T_t 表示处理的前期或后期; δ_i 表示个体固定效应; γ_t 表示时间固定效应。

若模型(2.3)设定正确,并且满足平行趋势假设,则估计 β_1 就可得到平均处理效应的估计。但在实际应用场景中,平行趋势假设过于严格,处理时间不同存在交错的情况等一系列问题。

2.2.2 合成控制法

合成控制法主要改进了双重差分法存在随时间变化的不可观测的混杂而引起的内生性问题,区别在于合成控制法中处理组只有一个个体。其主要思想是通过控制组中不同个体加权合成一个虚拟的与处理组相似的新的控制组,比较处理前后的处理组和新控制组的结果变量的变化,从而得出因果效应。

合成控制法假设个体 i 在时间 t 未受处理时的潜在结果是由可观测的协变量和不可观测的共同因子决定:

$$Y_{it} = \alpha_i + \beta_i^\top X_i + f_t^\top \lambda_i + \varepsilon_{it}, \quad (i = 1, \dots, N; t = 1, \dots, T), \quad (2.4)$$

其中, α_i 表示每个个体上载荷都相同的未知的共同因子; X_i 表示个体 i 的不受处理影响的观测协变量; β_i 表示随时间变化的未知参数; f_t 表示不可观测的但可随时间变化的共同因子; λ_i 表示个体 i 的未知的因子载荷; ε_{it} 为零均值的随机误差项。

2.2.3 因子估计方法

合成控制法和因子估计方法虽然都涉及因子结构，但在实践中有所不同。合成控制法并不直接估计因子模型，而是通过控制组因变量的加权组合来推断处理组的反事实结果。相反，因子估计方法则需要估计因子模型中的因子数量、因子和载荷参数，才能估计因变量的预测值来构建反事实。

假设未受处理的潜在结果由如下因子模型生成：

$$Y_{it} = \lambda_i^\top f_t + \varepsilon_{it}, \quad i=1, \dots, N; t=1, \dots, T, \quad (2.5)$$

其中， λ_i 表示个体 i 的因子载荷； f_t 表示 $r \times 1$ 的因子向量； ε_{it} 为零均值的随机误差项。定义 $F = (f_1, \dots, f_T)^\top$ ， $\Lambda = (\lambda_1, \dots, \lambda_N)^\top$ 。在式(2.5)中， F 和 Λ 不能同时识别，因此在估计时添加约束条件 $F^\top F = I_r$ 和 $\Lambda^\top \Lambda / N$ 是一个对角矩阵。当 N 固定且较小、误差 ε_{it} 服从正态分布时，传统因子分析法通常采用极大似然法估计式(2.5)。

在因子估计方法中都假定了因子个数 r 是已知的，然而在实际应用中因子个数是未知的。对于同一组数据，不同的因子个数估计方法可能估计出不同的因子个数，从而对因子模型的估计结果产生影响。因此需要对因子模型做更多的稳健性检验。

2.3 基于机器学习的因果推断方法

随着机器学习功能变得越来越强大，越来越多的机器学习算法融入因果推断方法中，常用弹性网络和核范数最小化矩阵填充方法。

2.3.1 弹性网络

弹性网络是一种用于线性回归和机器学习的正则化方法。弹性网络结合了 L_1 正则化和 L_2 正则化。 L_1 正则化即 Lasso，通过添加权重的绝对值到目标函数中，使模型参数稀疏化，即使只有少数特征对目标变量有显著影响。这有助于特征选择和模型解释性，但可能导致一些参数的值变为零，从而丢失了一些信息。 L_2 正

则化即岭回归，通过添加权重的平方和到目标函数中，惩罚大的参数值，从而减少了参数的过度波动，提高了模型的稳定性和泛化能力。

弹性网络的目标函数由拟合误差和正则化项两部分组成，如式(2.5)所示：

$$\hat{\mathbf{L}} = \min_{\mathbf{L}} \|\mathbf{Y} - \mathbf{L}\|_F^2 + \lambda_1 \|\mathbf{L}\|_1 + \lambda_2 \|\mathbf{L}\|_2, \quad (2.5)$$

其中， \mathbf{Y} 是特征矩阵含有缺失值； \mathbf{L} 为目标填充矩阵； $\|\mathbf{L}\|_1 = \sum_{i,j} |L_{ij}|$ ， $\|\mathbf{L}\|_2 = \sum_{i,j} |L_{ij}|^2$ ； λ_1 和 λ_2 为正则化参数， λ_1 控制稀疏性， λ_2 控制平滑性。

2.3.2 核范数矩阵填充方法

Athey 等(2021)提出可利用核范数矩阵填充方法估计缺失潜在结果，将矩阵填充方法与因果推断方法相融合。矩阵填充是利用矩阵奇异值的稀疏性，即矩阵的低秩性，通过凸优化问题最小化核范数，精确地填充缺失元素的方法。核范数最小化低秩矩阵逼近问题解决：

$$\min_{\mathbf{L}} \frac{1}{|\Omega|} \|P_{\Omega}(\mathbf{Y} - \mathbf{L})\|_F^2 + \lambda \|\mathbf{L}\|_*, \quad (2.6)$$

式中， $\mathbf{Y} \in R^{N \times T}$ 为观测矩阵， $\mathbf{L} \in R^{N \times T}$ 为目标恢复矩阵， $\|\cdot\|_*$ 为矩阵的核范数， λ 为调节参数， Ω 为已知观测元素的位置， P_{Ω} 为正交投影算子。

问题(2.6)可利用 Cai 等(2010)提出的奇异值阈值算法(Singular Value Thresholding, SVT)快速求解。在 SVT 算法中，需要求解奇异值阈值算子：

$$S_{\lambda}(\mathbf{Y}) = \min_{\mathbf{L} \in R^{N \times T}} \|\mathbf{Y} - \mathbf{L}\|_F^2 + \lambda \|\mathbf{L}\|_*, \quad (2.7)$$

该算子的求解需要如下引理：

引理 1 记 $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\top}$ 是 \mathbf{Y} 的奇异值分解，问题(2.6)的最优解为

$$\mathbf{L}^* = S_{\lambda}(\mathbf{Y}) = \mathbf{U}S_{\lambda}(\mathbf{\Sigma})\mathbf{V}^{\top} \quad (2.8)$$

其中 $S_{\lambda}(\mathbf{\Sigma})$ 带参数 λ 的对角矩阵 $\mathbf{\Sigma}$ 的软阈值函数， $S_{\lambda}(\mathbf{\Sigma})_{ii} = \max(\Sigma_{ii} - \lambda, 0)$ 。

2.4 小结

本章主要介绍了反事实框架和因果推断方法。首先介绍了反事实框架包含的

反事实结果、重要假设以及在实际应用中所关注的因果效应。其次介绍了在反事实框架下的经典因果推断方法中的 DID、SCM 和因子估计方法。最后介绍了机器学习与因果推断相融合的方法弹性网络和核范数矩阵填充方法。在后续研究中提出的方法与这几种方法进行对比；通过了解反事实框架和常用的因果推断方法，能更深入地掌握当前研究现状，并对其中存在的问题进行全面总结。

3 基于加权核范数最小化的矩阵填充反事实结果估计

3.1 引言

在因果推断领域,随机对照试验通常被视为确定因果效应的黄金标准。然而,在实际应用中进行随机试验往往是困难甚至不可行的。因此,研究人员转向观测数据。在观察性研究中,由于无法同时观察到个体的两种潜在结果,传统方法大多通过“构造”与观测样本特征相似的反事实样本,以比较与实际观测值的差异,并由此估计因果效应。然而,这种方法可能受到过多的假设的限制,从而影响结果的准确性。Imbens 等学者将机器学习方法引入到因果推断分析框架中。因果效应推断的核心挑战在于总是存在潜在结果缺失的问题。为了克服这一问题,利用机器学习方法通过求解约束优化问题填补矩阵缺失值,减少传统因果推断所依赖的假设。这样做可以更精确地估计不可观测的潜在结果,从而实现了对因果效应的估计和推断。这一方法已成为各界学者研究和关注的重点。

矩阵填充在机器学习、推荐系统、图像恢复以及视频去噪等领域都有着广泛的应用。矩阵填充问题一般在给定部分观察数据的矩阵,估计矩阵缺失元素。矩阵填充通常考虑随机缺失数据的模式,然而在因果推断应用中数据缺失来自于干预的选择,并且缺失数据有特定结构,如块状、条状缺失、交错缺失。矩阵填充对条块缺失数据修复有其独特优势。机器学习方法中已有大量矩阵填充方法的应用,这些估计量依赖于“正则化”考虑更加复杂的模型,特别是核范数正则化。Candès 和 Recht(2009)首次提出用核范数作为惩罚函数代替秩函数来解决矩阵填充问题,会使得整个优化问题变得更容易计算,不再是 NP-难。并假设完整矩阵是低秩矩阵加上噪声,以及缺失是完全随机的。后续大部分研究均在此基础上放松约束条件再最小化目标矩阵的核范数或预测值与实际观测值差异的平方和。然而,对目标矩阵中所有奇异值使用同一值收缩,使得核范数正则化方法过度惩罚较大奇异值,得到的解偏离真实解。Gu 等(2014)提出加权核范数矩阵填充方法,通过构造自适应权重提高核范数的灵活性。张玮奇等(2015)给出了加权核范数最小化模型可用收缩算子得到全局最优解。石莹等(2022)基于软阈值算法思想引入不精确近邻算子提出加权核范数松弛原始低秩极小化问题,降低算法的时间复杂度。

在反事实框架下，因果推断问题转换为反事实估计问题，Abadie 等(2010)提出利用合成控制法进行比较案例研究，通过控制组特征合成处理组的反事实结果。Athey 等(2019)利用机器学习集成方法预测反事实结果比传统方法更精确。在因子分析视角下，Bai 等(2021)和 Xiong 等(2023)提出带有特殊结构的因子模型，通过估计因子和载荷得到反事实结果。而 Athey 等(2021)借鉴关于因子模型、交互固定效应以及矩阵填充的文献，从一个不完全观测的矩阵出发，提出求解核范数正则化优化问题填补缺失元素，即反事实值，而不是一致的估计因子和载荷。其重要贡献之一是论证了矩阵填充回、合成控制回归、弹性网估计和双重差分回归估计量可以看作相同目标函数，只是在正则化和参数限制有所不同。

本章基于面板数据，综合考虑个体和时间两个维度地固定效应，将加权核范数矩阵填充模型与因果推断相融合估计反事实结果，加权核范数矩阵填充算法依赖奇异值分解，对较小的奇异值进行较多的惩罚，对较大的奇异值进行较小的惩罚，尽可能保留数据主要信息，并利用 ADMM 框架优化问题。最后，在模拟数据集和真实数据集上验证所提模型的可行性。

3.2 理论基础

矩阵填充问题一般在给定部分观察数据的矩阵，估计矩阵缺失元素。Athey 等(2021)提出核范数最小化矩阵填充方法估计因果效应，其中设定 $\mathbf{Y}(0)$ 和 $\mathbf{Y}(1)$ 均是 $N \times T$ 的矩阵，定义 Ω 为对应 \mathbf{Y} 中 $W_{it} = 1$ 对应的 (i, t) 中的集合项，

$$\mathbf{Y}(0) = \begin{pmatrix} Y_{11} & Y_{12} & ? & \dots & Y_{1T} \\ ? & ? & Y_{23} & \dots & ? \\ Y_{31} & ? & Y_{33} & \dots & ? \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{N1} & ? & Y_{N3} & \dots & ? \end{pmatrix}, \quad \mathbf{Y}(1) = \begin{pmatrix} ? & ? & Y_{13} & \dots & ? \\ Y_{21} & Y_{22} & ? & \dots & Y_{2T} \\ ? & Y_{32} & ? & \dots & Y_{3T} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ? & Y_{N2} & ? & \dots & Y_{NT} \end{pmatrix},$$

其中， $\mathbf{Y}(0)$ 表示没有受到干预的控制组，即 $W_{it} = 0$ ； $\mathbf{Y}(1)$ 表示受到干预的干预组，即 $W_{it} = 1$ ；

$$W_{it} = \begin{cases} 1 & (i, j) \in \Omega \\ 0 & (i, j) \notin \Omega \end{cases},$$

其中出现在 $\mathbf{Y}(1)$ 的数据，说明个体受到干预，所以不能观测它在 $\mathbf{Y}(0)$ 的数据，

因此在 $\mathbf{Y}(0)$ 对应的数据是一个问号“?”，表示数据缺失。如果要填充这些缺失数据，可以通过观察其他比较相同的个体，通过他们的特征观测问号的值。通过相似的对照补全信息，这个过程就是矩阵填充。其主要工作就是找到关于在 \mathbf{Y} 中缺失值的统计问题(即矩阵中的“?”的估计值)，缺失值表示干预后的观测结果的反事实。一旦这些值被输入，就可以估计平均因果效应 τ 。因此，考虑观测矩阵 \mathbf{Y} ，其期望值矩阵为 \mathbf{L} ，则观测矩阵可以表示为：

$$Y_{it} = \begin{cases} L_{it} + \varepsilon_{it} & W_{it} = 1 \\ 0 & W_{it} = 0 \end{cases},$$

其中 \mathbf{Y} 是 $N \times T$ 的观测矩阵，目标是要估计 \mathbf{L} 是要估计的完整矩阵， ε 是独立于 \mathbf{L} 的噪声矩阵。

为了方便起见， $\mathbb{R}^{n_1 \times n_2}$ 表示 $n_1 \times n_2$ 的实矩阵全体， $\text{rank}(\mathbf{X})$ 表示矩阵 \mathbf{X} 的秩， X_{ij} 表示矩阵 \mathbf{X} 位于 (i, j) 的元素， $\|\mathbf{X}\|_F$ ， $\|\mathbf{X}\|_*$ 分别表示矩阵 \mathbf{X} 的 Frobenius 范数和核范数，用 $\langle \mathbf{X}, \mathbf{Y} \rangle = \text{trace}(\mathbf{X}^\top \mathbf{Y})$ 表示矩阵 \mathbf{X} 和矩阵 \mathbf{Y} 的内积(即有 $\|\mathbf{X}\|_F^2 = \langle \mathbf{X}, \mathbf{X} \rangle$)， Ω 为 $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ 的子集，表示已知的观察元素的位置， $P_\Omega(\cdot)$ 表示正交投影算子定义为：

$$P_\Omega(\mathbf{M}) = \begin{cases} M_{ij} & (i, j) \in \Omega \\ 0 & (i, j) \notin \Omega \end{cases}, \text{ 和 } P_\Omega^\perp(\mathbf{M}) = \begin{cases} 0 & (i, j) \in \Omega \\ M_{ij} & (i, j) \notin \Omega \end{cases}.$$

3.3 加权核范数矩阵填充反事实结果估计

3.3.1 加权核范数最小化模型

核范数最小化在低秩矩阵逼近中得到了广泛的应用，但仍存在一些问题。为了确保凸性，标准核范数对每个奇异值进行同等处理，使得式(2.8)中的软阈值算子以相同的 λ 收缩每个奇异值。

由于不同的奇异值具有不同的重要程度，因此对每个奇异值赋予不同的权重更贴近实际情况。因此，为了提高核范数的灵活性，提出使用加权核范数，并研究其最小化问题。矩阵 \mathbf{L} 的加权核范数定义为：

$$\|\mathbf{L}\|_{w,*} = \sum_{i=1}^r w_i \sigma_i(\mathbf{L}), \quad (3.1)$$

其中 $\mathbf{w} = [w_1, w_2, \dots, w_r]^\top$ 且 $w_i \geq 0$ 为分配给 $\sigma_i(\mathbf{L})$ 的非负权重。

通过引入权重向量, 加权核范数模型提高了原始核范数模型的灵活性。然而, 权重的设置对模型起着至关重要的作用。借鉴 Gu 等(2017)提出的通过自适应加权来增强稀疏性的思想, 对矩阵奇异值进行自适应加权调整:

$$w_i^{l+1} = \frac{C}{\sigma_i(L_l) + \varepsilon},$$

其中 $\sigma_i L_l$ 表示第 l 次迭代时的奇异值; w_i^{l+1} 表示第 l 次迭代第 i 个奇异值的权重; ε 为小的正数, 防止除以 0; C 为常数。

通过使用 Frobenius 范数来度量观测数据矩阵 \mathbf{Y} 和潜在数据矩阵 \mathbf{L} 之间的差异, 并考虑对较小的奇异值进行更多的惩罚, 对较大的奇异值进行更小的惩罚, 将(2.6)转换为(3.2),

$$\min_{\mathbf{L}, \mathbf{\Gamma}, \mathbf{\Delta}} \frac{1}{|\Omega|} \left\| P_{\Omega}(\mathbf{Y} - \mathbf{L} - \mathbf{\Gamma} \mathbf{I}_T^\top - \mathbf{I}_N \mathbf{\Delta}^\top) \right\|_F^2 + \lambda \|\mathbf{L}\|_{w,*}. \quad (3.2)$$

模型(3.2)通过引入变量 $\mathbf{\Gamma} \in R^{N \times 1}$ 和 $\mathbf{\Delta} \in R^{T \times 1}$ 来估计 \mathbf{L} 中包含的个体固定效应和时间固定效应, 然而, 由于模型(3.2)中引入加权核范数使得目标函数是非凸且非光滑, 次梯度条件不在满足, 因此加权核范数正则化问题比核范数正则化更难优化。

3.3.2 最优解证明

由于矩阵 \mathbf{L} 中包含固定效应, 则 $\mathbf{L}^* = \hat{\mathbf{L}} + \hat{\mathbf{\Gamma}} \mathbf{I}_N^\top + \mathbf{I}_N \hat{\mathbf{\Delta}}^\top$ 。为了方便论证最优解, 首先假设 \mathbf{L} 中不包含固定效应, 因此不需要估计 $\mathbf{\Gamma}$ 和 $\mathbf{\Delta}$ 。加权核范数正则化依赖奇异值分解, 由于奇异值在矩阵 $\mathbf{\Sigma}$ 中是按降序排列的, 为了确保较大的奇异值用较小值收缩, 权重 \mathbf{w} 需要满足 $0 \leq w_1 \leq \dots \leq w_n$, 即 \mathbf{w} 为升序排列。如果权重满足 $0 \leq w_1 \leq \dots \leq w_n$, 可得到全局最优解。

引理 2 对任意给定矩阵 $\mathbf{A} \in R^{n \times n}$ 和非负对角矩阵 $\mathbf{W} \in R^{n \times n}$, 其中对角元素是

降序排列的，且令 $\mathbf{A} = \mathbf{X}\Phi\mathbf{Y}^\top$ 是 \mathbf{A} 的奇异值分解，则有

$$\sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{W}) = \max_{\mathbf{U}^\top\mathbf{U}=\mathbf{I}, \mathbf{V}^\top\mathbf{V}=\mathbf{I}} \text{tr}[\mathbf{W}\mathbf{U}^\top\mathbf{A}\mathbf{V}], \quad (3.3)$$

其中， \mathbf{I} 是单位阵， $\sigma_i(\mathbf{A})$ 、 $\sigma_i(\mathbf{W})$ 分别是 \mathbf{A} 和 \mathbf{W} 的奇异值。当 $\mathbf{U} = \mathbf{X}$ 、 $\mathbf{V} = \mathbf{Y}$ 时， $\text{tr}[\mathbf{W}\mathbf{U}^\top\mathbf{A}\mathbf{V}]$ 达到最大值。

定理 1 对任意给定矩阵 $\mathbf{Y} \in \mathbb{R}^{N \times T}$ 且其奇异值分解为 $\mathbf{Y} = \mathbf{U}\Sigma\mathbf{V}^\top$ ，其中 $\Sigma = \begin{pmatrix} \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r) \\ 0 \end{pmatrix} \in \mathbb{R}^{N \times T}$ 。则加权核范数问题的解可表示为：

$$\mathbf{L}^* = \mathbf{U}\hat{\mathbf{D}}\mathbf{V}^\top,$$

其中 $\hat{\mathbf{D}} = \begin{pmatrix} \text{diag}(d_1, d_2, \dots, d_r) \\ 0 \end{pmatrix}$ 是非负对角矩阵，并且 (d_1, d_2, \dots, d_r) 是如下凸优化问题的最优解：

$$\begin{aligned} \min_{d_1, d_2, \dots, d_r} \sum_{i=1}^r (\sigma_i - d_i)^2 + w_i d_i \\ \text{s.t.} \quad d_1 \geq d_2 \geq \dots \geq d_r \geq 0 \end{aligned} \quad (3.4)$$

证明：对于任何 $\mathbf{L} \in \mathbb{R}^{N \times T}$ ，其 SVD 可以表示为 $\mathbf{L} = \bar{\mathbf{U}}\mathbf{D}\bar{\mathbf{V}}^\top$ ，其中 $\bar{\mathbf{U}}$ 和 $\bar{\mathbf{V}}$ 是酉矩阵，并且 $\mathbf{D} = \begin{pmatrix} \text{diag}(d_1, d_2, \dots, d_r) \\ 0 \end{pmatrix}$ 与 $d_1 \geq d_2 \geq \dots \geq d_r \geq 0$ 。那么有

$$\begin{aligned} & \min_{\mathbf{L}} \|\mathbf{Y} - \mathbf{L}\|_F^2 + \|\mathbf{L}\|_{w,*} \\ \Leftrightarrow & \min_{\bar{\mathbf{U}}, \bar{\mathbf{V}}, \mathbf{D}} \|\mathbf{Y} - \bar{\mathbf{U}}\mathbf{D}\bar{\mathbf{V}}^\top\|_F^2 + \|\bar{\mathbf{U}}\mathbf{D}\bar{\mathbf{V}}^\top\|_{w,*} \\ \Leftrightarrow & \min_{\bar{\mathbf{U}}, \bar{\mathbf{V}}, \mathbf{D}} \|\mathbf{Y}\|_F^2 - 2\text{tr}(\mathbf{Y}\bar{\mathbf{U}}^\top\mathbf{D}\bar{\mathbf{V}}) + \|\mathbf{D}\|_F^2 + \|\mathbf{D}\|_{w,*} \\ \Leftrightarrow & \min_{\mathbf{D}} \|\mathbf{D}\|_F^2 + \|\mathbf{D}\|_{w,*} - 2 \max_{\bar{\mathbf{U}}^\top\bar{\mathbf{U}}=\mathbf{I}, \bar{\mathbf{V}}^\top\bar{\mathbf{V}}=\mathbf{I}} \text{tr}(\mathbf{Y}\bar{\mathbf{U}}^\top\mathbf{D}\bar{\mathbf{V}}) \end{aligned}$$

根据引理 2，有

$$\max_{\bar{\mathbf{U}}^\top\bar{\mathbf{U}}=\mathbf{I}, \bar{\mathbf{V}}^\top\bar{\mathbf{V}}=\mathbf{I}} \text{tr}[\mathbf{Y}\bar{\mathbf{U}}^\top\mathbf{D}\bar{\mathbf{V}}] = \sum_i \sigma_i d_i,$$

当 $\bar{\mathbf{U}} = \mathbf{U}$ ， $\bar{\mathbf{V}} = \mathbf{V}$ 时有最优解。因此

$$\begin{aligned}
& \min_L \|Y - L\|_F^2 + \|L\|_{w,*} \\
& \Leftrightarrow \min_D \|D\|_F^2 + \|D\|_{w,*} - 2\sum_{i=1}^r d_i \sigma_i \\
& \Leftrightarrow \min_D \sum_{i=1}^r d_i^2 + w_i d_i - 2d_i \sigma_i \\
& \Leftrightarrow \min_D \sum_{i=1}^r (d_i - \sigma_i)^2 + w_i d_i
\end{aligned}$$

从上述推导过程可以看出加权核范数矩阵填充问题(3.2)的最优解为:

$$L^* = UDV^\top, \quad (3.5)$$

其中 D 是优化问题(3.4)的最优解, 证毕。

定理 1 表明, 非凸的加权核范数最小化问题总可以通过(3.4)得到全局最优解。下面推论更进一步证明当权重为升序排列时, 仍具有全局最优闭式解。

推论 1 当 $0 \leq w_1 \leq \dots \leq w_n$ 时, 问题(3.2)的全局最优解为

$$D^* = \begin{pmatrix} \text{diag}(\bar{d}_1, \bar{d}_2, \dots, \bar{d}_r) \\ 0 \end{pmatrix}, \quad \text{其中 } \bar{d}_i = \max\left(\sigma_i - \frac{w_i}{2}, 0\right).$$

证明: 如果忽略问题(3.4)中的约束条件, 可以得到无约束条件问题:

$$\min_{d_i \geq 0} (\sigma_i - d_i)^2 + w_i d_i, \quad (3.6)$$

当 $d_i \geq 0$, $i = 1, 2, \dots, r$, 上述问题等价于下面的问题:

$$\min_{d_i \geq 0} \left(d_i - \left(\sigma_i - \frac{w_i}{2} \right) \right)^2,$$

很容易可以看出问题(3.6)的全局最优解为:

$$\bar{d}_i = \max\left(\sigma_i - \frac{w_i}{2}, 0\right), \quad i = 1, 2, \dots, n.$$

当 $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n$, $w_1 \leq w_2 \leq \dots \leq w_n$ 时, 则 $\bar{d}_1 \geq \bar{d}_2 \geq \dots \geq \bar{d}_n \geq 0$ 。因此, 当 $\bar{d}_{i=1,2,\dots,n}$ 满足约束条件时, \bar{d}_i 为问题(3.4)的全局最优解, 证毕。

矩阵奇异值按降序排列, 较大的奇异值携带更多重要信息, 因此需要缩小较大的奇异值, 以保持数据主要特征。推论 1 保证了加权核范数矩阵填充问题的最优解为加权奇异值软阈值算子:

$$S_{\lambda \| \cdot \|_{w,*}}(Y) = US_{w/2}(\Sigma)V^\top,$$

其中 \mathbf{Y} 的奇异值分解为 $\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, $S_{w/2}(\mathbf{\Sigma})$ 是权重为 \mathbf{w} 的广义软阈值算子,
 $S_{w/2}(\mathbf{\Sigma})_{ii} = \max(\Sigma_{ii} - w_i/2, 0)$ 。

3.3.3 模型优化

为求解模型(3.2), 引入辅助变量 $\mathbf{H} \in \mathbb{R}^{N \times T}$ 来放松约束, 使用 ADMM 交替乘子法等效优化问题求解, 则有:

$$\begin{aligned} \min_{L, \mathbf{A}} \frac{1}{|\Omega|} & \left\| P_{\Omega}(\mathbf{Y} - \mathbf{L} - \mathbf{G}\mathbf{I}_T^\top - \mathbf{I}_N \mathbf{A}^\top) \right\|_F^2 + \lambda \|\mathbf{L}\|_{w,*} \\ \text{s.t.} \quad & \mathbf{L} = \mathbf{H} \end{aligned} \quad (3.7)$$

由于固定效应只出现在平方项中, 为了方便优化, 忽略模型(3.7)中的固定效应, 则构建增广拉格朗日函数为:

$$\mathcal{L}(\mathbf{H}, \mathbf{L}, \mathbf{A}) = \left\| P_{\Omega}(\mathbf{Y} - \mathbf{L}) \right\|_F^2 + \lambda \|\mathbf{L}\|_{w,*} + \text{tr}[\mathbf{A}^\top (\mathbf{L} - \mathbf{H})] + \frac{\mu}{2} \|\mathbf{L} - \mathbf{H}\|_F^2, \quad (3.8)$$

其中 \mathbf{A} 是拉格朗日乘子, $\mu > 0$ 是罚参数, $\text{tr}(\cdot)$ 表示矩阵的迹, 如 $\text{tr}(\mathbf{M}) = \sum_{i=1}^n m_{ii}$ 。

在 k 次迭代中, 需要交替计算 \mathbf{H}_{k+1} 、 \mathbf{L}_{k+1} 和 \mathbf{A}_{k+1} 。首先, 固定 \mathbf{L}_k 和 \mathbf{A}_k 更新 \mathbf{H}_{k+1} ,

$$\begin{aligned} \mathbf{H}_{k+1} &= \arg \min_{\mathbf{H}} \mathcal{L}(\mathbf{H}, \mathbf{L}_k, \mathbf{A}_k) \\ &= \arg \min_{\mathbf{H}} \text{tr}[\mathbf{A}_k^\top (\mathbf{L}_k - \mathbf{H})] + \frac{\mu_k}{2} \|\mathbf{L}_k - \mathbf{H}\|_F^2 \\ &= \arg \min_{\mathbf{H}} \frac{\mu_k}{2} \left\| \mathbf{H} - \left(\mathbf{L}_k + \frac{1}{\mu_k} \mathbf{A}_k \right) \right\|_F^2 \\ &= \mathbf{L}_k + \frac{1}{\mu_k} \mathbf{A}_k \end{aligned}$$

其次, 固定 \mathbf{H}_{k+1} 和 \mathbf{A}_k 更新 \mathbf{L}_{k+1} ,

$$\begin{aligned} \mathbf{L}_{k+1} &= \arg \min_{\mathbf{L}} \mathcal{L}(\mathbf{H}_{k+1}, \mathbf{L}, \mathbf{A}_k) \\ &= \arg \min_{\mathbf{L}} \lambda \|\mathbf{L}\|_{w,*} + \text{tr}[\mathbf{A}_k^\top (\mathbf{L} - \mathbf{H}_{k+1})] + \frac{\mu_k}{2} \|\mathbf{L} - \mathbf{H}_{k+1}\|_F^2 \\ &= \arg \min_{\mathbf{L}} \lambda \|\mathbf{L}\|_{w,*} + \frac{\mu_k}{2} \left\| \mathbf{L} - \left(\mathbf{H}_{k+1} - \frac{1}{\mu_k} \mathbf{A}_k \right) \right\|_F^2 \\ &= S_{\lambda/\mu_k} \left(\mathbf{H}_{k+1} - \frac{1}{\mu_k} \mathbf{A}_k \right) \end{aligned}$$

其中 $S_\lambda(\mathbf{L})$ 是加权核范数奇异值收缩算子。再次，固定 H_{k+1} 和 L_{k+1} ，更新 A_{k+1} ，

$$A_{k+1} = A_k + \mu_k (L_{k+1} - H_{k+1}),$$

然后，更新参数 μ_k ：

$$\mu_{k+1} = \rho * \mu_k,$$

其中 $\mu_0 > 0$ ， $\rho > 1$ ， $k = 1, 2, \dots$ 。

通过交叉验证选择最佳参数 λ ，直到序列 $\{L_k(\lambda, \Omega)\}$ 收敛。在每次迭代得到 L_{k+1} 后，使用一阶条件估计 Γ_{k+1} 和 Δ_{k+1} 。最后，经过多轮迭代更新，如果

$$\|\mathbf{Y} - L_{k+1}\|_F / \|\mathbf{Y}\|_F \leq \varepsilon,$$

其中 $\varepsilon = 10^{-6}$ ，则判断终止迭代，由此得到加权核范数矩阵填充算法估计的潜在结果并得到 RMSE。

根据上述推导过程将加权核范数矩阵填充算法总结为：

算法 1 加权核范数矩阵填充算法

输入：缺失矩阵 \mathbf{Y} ，位置矩阵 Ω ，

输出： $\mathbf{L} = L_{k+1}$ 。

step1: 初始化： $\mu_0 > 0$ ， $\rho > 1$ ， $\varepsilon = 10^{-6}$ ， $k = 0$ ， $L_0 = \mathbf{Y}$ ， $H_0 = 0$ ；

step 2: for $k = 1, 2, \dots, T$ do

step 3: $H_{k+1} = L_k + \frac{1}{\mu_k} A_k$ ；

step 4: $L_{k+1} = S_{1/\mu_k} \left(H_{k+1} - \frac{1}{\mu_k} A_k \right)$ ；

step 5: $A_{k+1} = A_k + \mu_k (L_{k+1} - H_{k+1})$ ；

step 6: 更新 $\mu_{k+1} = \rho * \mu_k$ ；

step 7: while $\|\mathbf{Y} - L_{k+1}\|_F / \|\mathbf{Y}\|_F \leq \varepsilon$ ；

step 8: end for

在算法 1 的迭代过程中, 得到的解 L_k 不断地逼近真实解, 并且目标函数是单调递减的, 直到收敛到某个稳定点。下面定理 2 保证了算法 1 的收敛性。

定理 2 若权重按照升序排列, 由算法 1 生成的序列 L_k 满足

$$\lim_{k \rightarrow \infty} \|L_{k+1} - L_k\|_F = 0.$$

证明: 记第 $k+1$ 次迭代时 $\left\{H_k - \frac{1}{\mu_k} A_k\right\}$ 的奇异值分解为 $U_k \Lambda_k V_k^\top$, 其中

$$\Lambda_k = \left\{diag(\sigma_k^1, \sigma_k^2, \dots, \sigma_k^n)\right\}, \text{ 由定理 1 知: } L_{k+1} = U_k \Sigma_k V_k^\top, \text{ 其中 } \Sigma_k = S_{w/\mu_k}(\Lambda_k).$$

基于算法 1 步骤 5 中的拉格朗日乘子更新方法可得:

$$\begin{aligned} \|A_{k+1}\|_F &= \|A_k + \mu_k (L_{k+1} - H_{k+1})\|_F \\ &= \mu_k \|\mu_k^{-1} A_k + L_{k+1} - H_{k+1}\|_F \\ &= \mu_k \|U_k \Lambda_k V_k^\top - U_k \Sigma_k V_k^\top\|_F \\ &= \mu_k \|\Lambda_k - \Sigma_k\|_F \\ &= \mu_k \left\| \Lambda_k - S_{\frac{w}{\mu_k}}(\Lambda_k) \right\|_F \\ &\leq \mu_k \sqrt{\sum_i \left(\frac{w_i}{\mu_k}\right)^2} \\ &= \sqrt{\sum_i w_i^2} \end{aligned}$$

因此, $\{A_k\}$ 是有界的。为分析 $\mathcal{L}(H, X, A)$ 的有界性, 在迭代中下面不等式成立:

$$\mathcal{L}(H_{k+1}, L_{k+1}, A_k) \leq \mathcal{L}(H_k, L_k, A_k),$$

然后更新 A : $A_{k+1} = A_k + \mu_k (L_{k+1} - H_{k+1})$ 。

$$\begin{aligned} &\mathcal{L}(H_k, L_k, A_k) \\ &= \mathcal{L}(H_k, L_k, A_{k-1}) + \frac{\mu_k - \mu_{k-1}}{2} \|L_k - H_k\|_F^2 + \langle A_k - A_{k-1}, L_k - H_k \rangle \\ &= \mathcal{L}(H_k, L_k, A_{k-1}) + \frac{\mu_k - \mu_{k-1}}{2} \|\mu_{k-1}^{-1} (A_k - A_{k-1})\|_F^2 + \langle A_k - A_{k-1}, \mu_{k-1}^{-1} (A_k - A_{k-1}) \rangle \\ &= \mathcal{L}(H_k, L_k, A_{k-1}) + \frac{\mu_k - \mu_{k-1}}{2\mu_{k-1}^2} \|A_k - A_{k-1}\|_F^2 \end{aligned}$$

当 $k=1, \dots, \infty$ 时, 记 Θ 为 $\|A_k - A_{k-1}\|_F^2$ 的上界, 则

$$\mathcal{L}(H_{k+1}, L_{k+1}, A_k) \leq \mathcal{L}(H_1, L_1, A_0) + \Theta \sum_{k=1}^{\infty} \frac{\mu_k - \mu_{k-1}}{2\mu_{k-1}^2},$$

当 $\{\mu_k\}$ 满足 $\sum_{k=1}^{\infty} \mu_k^{-2} \mu_{k+1} < +\infty$ 时, 则 $\sum_{k=1}^{\infty} \frac{\mu_k - \mu_{k-1}}{2\mu_{k-1}^2} \leq \sum_{k=1}^{\infty} \mu_{k-1}^{-2} \mu_k < +\infty$ 。因此,

$\mathcal{L}(H_k, L_k, A_k)$ 有上界, 则至少存在一个聚点为目标函数的可行解。

由定理 1 及算法 1 可知:

$$L_k = U_{k-1} S_{w/\mu_{k-1}} (\Lambda_{k-1}) V_{k-1}^\top, \quad L_{k+1} = H_{k+1} + \mu_k^{-1} A_{k+1} - \mu_k^{-1} A_k,$$

其中 $U_{k-1} \Lambda_{k-1} V_{k-1}^\top$ 为 $\left\{ H_k - \frac{1}{\mu_k} A_k \right\}$ 第 k 次迭代时的奇异值分解。那么

$$\begin{aligned} & \lim_{k \rightarrow \infty} \|L_{k+1} - L_k\|_F \\ &= \lim_{k \rightarrow \infty} \left\| \left(H_{k+1} + \mu_k^{-1} A_{k+1} - \mu_k^{-1} A_k \right) - L_k \right\|_F \\ &= \lim_{k \rightarrow \infty} \left\| \left(H_{k+1} + \mu_k^{-1} A_{k+1} - \mu_k^{-1} A_k \right) - L_k + \mu_{k-1}^{-1} A_{k-1} - \mu_{k-1}^{-1} A_{k-1} \right\|_F \\ &\leq \lim_{k \rightarrow \infty} \left\| H_{k+1} - \mu_{k-1}^{-1} A_{k-1} - L_k \right\|_F + \left\| \mu_k^{-1} A_{k+1} - \mu_k^{-1} A_k + \mu_{k-1}^{-1} A_{k-1} \right\|_F \\ &= \lim_{k \rightarrow \infty} \left\| \Lambda_{k-1} - S_{w/\mu_{k-1}} (\Lambda_{k-1}) \right\|_F \\ &= 0 \end{aligned}$$

在定理 2 的基础上, 加权核范数最小化矩阵填充方法的收敛性通过目标函数逼近最优值来解释。

3.4 数值检验

3.4.1 模拟数据生成

为验证提出的加权核范数矩阵填充算法的有效性, 随机生成矩阵 $A \in R^{N \times R}$ 和 $B \in R^{R \times T}$, $\Gamma \in R^{N \times 1}$, $\Delta \in R^{T \times 1}$, 令 $L = AB + \Gamma \Gamma_T^\top + I_N \Delta^\top + N$, 则 $L \in R^{N \times R}$, 其中矩阵 A , B 和 N 中每个元素是独立的高斯随机变量, N 为方差为 0.1 的高斯噪声矩阵。

3.4.2 方法模拟与精度比较

在生成的模拟数据集上, 人为指定受处理的个体和时间点, 比较研究常用的

5 个估计量和加权核范数矩阵填充估计量:

- (1)DID: 双重差分法是基于个体和时间固定效应的回归推断;
- (2)SCM: 合成控制法为 Abadie 等(2010)最初提出的基于交互固定效应的有约束回归;
- (3)VT-EN: 弹性网络正则化的垂直回归, 放松了合成控制估计量的限制;
- (4)HR-EN: 弹性网络正则化的水平回归;
- (5)MC-NNM: Athey 等(2021)提出的核范数最小化矩阵填充;
- (6)MC-WNNM: 本章提出的基于加权核范数的矩阵填充方法。

对伪处理对应的预测值与实际值进行比较, 输出每个算法的均方根误差 (RMSE):

$$RMSE = \left\| L - \hat{L} \right\|_F / \sqrt{NT}.$$

在实验中, 参考 Athey 等(2021)中采用的随机同时处理和随机交错处理两种方式对不同规模矩阵进行验证, 参数 λ 由 K 折交叉验证选择最优值, 每个算法独立运行 10 次, 得到平均 RMSE。其中, 随机同时处理是随机选择个体 N_i 在 $T_0 + 1$ 期进行处理, 其余个体不接受处理; 随机交错处理是随机选择个体 N_i 在 T 期之后某个时期进行处理, 实际处理时间随个体随机变化。

表 3.1 不同算法在随机同时处理方式下 RMSE

	T_0	DID	SCM	VT-EN	HR-EN	MC-NNM	MC-WNNM
50×50	5	1.250785	1.889540	0.805120	1.234688	0.900086	0.850139
	15	1.230214	1.874975	0.649776	0.754214	0.260093	0.094119
	25	1.089687	1.882857	0.535731	0.656511	0.098357	0.055888
	35	0.978340	1.91752	0.559474	0.445741	0.060825	0.051827
	45	1.096089	1.922667	0.589541	0.539849	0.060785	0.057440
200×200	20	1.370007	2.363245	0.668859	0.985748	0.530285	0.153510
	60	1.425857	2.405319	0.696343	0.821004	0.336515	0.114383
	100	1.467124	2.381943	0.722731	0.819383	0.224616	0.109513
	140	1.444249	2.328814	0.805935	0.850235	0.167423	0.109685
	180	1.322852	2.308448	0.589413	0.570504	0.159079	0.111055

随机同时处理方式的实验结果如表 3.1 所示。从表 3.1 可以看出，在不同的处理时间 T_0 情况下，DID 和 SCM 的表现较差，这表明数据足够丰富时可以采用更复杂的模型。在 50×50 的数据集上，当 T_0/T 较小，即观测率较小时，HR-EN 表现较差；当 T_0/T 接近 1，即观测率较高时，MC-WNNM 算法表现良好。在 200×200 的数据集上，当 T_0/T 较小时，SCM 表现较差；当 T_0/T 接近 1 时，MC-NNM 和 MC-WNNM 算法都表现良好。

随机同时处理方式的误差棒图如下图 3.1 所示。

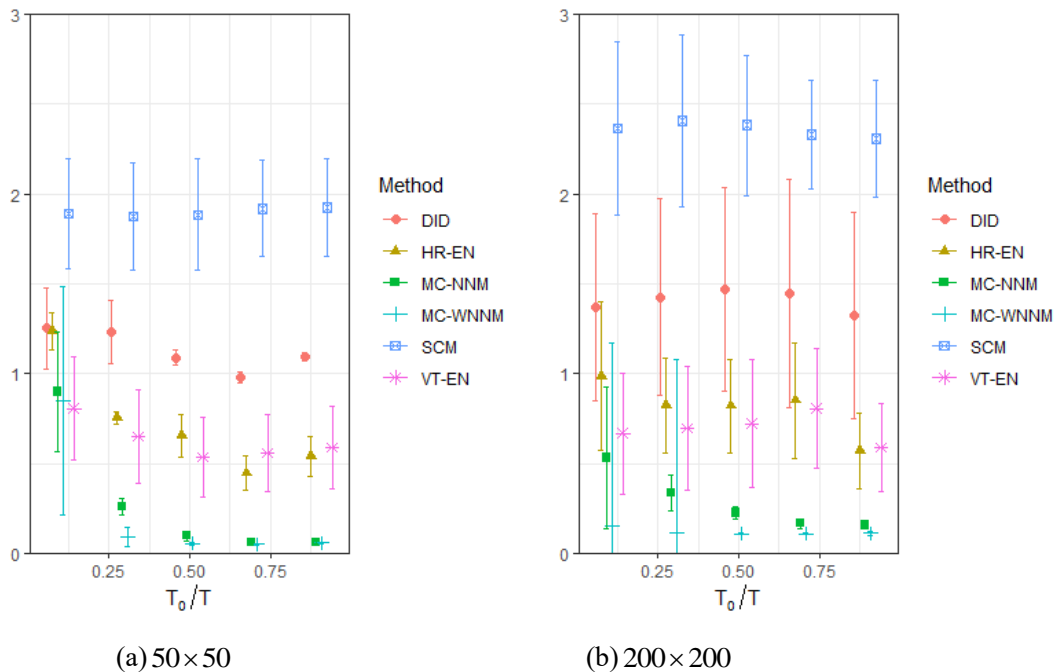


图 3.1 不同算法在随机同时处理方式下误差棒图

如图 3.1 可以看出，在随机同时处理方式下 SCM 和 DID 的误差大且不稳定，而矩阵填充方法 MC-NNM 和 MC-WNNM 的估计精度较高且误差区间小。当 MC-WNNM 算法在不同规模数据集上观测率大于 50% 时，MC-WNNM 算法估计精度明显优于其他算法。

如表 3.2 所示，随机交错处理方式的实验结果与随机同时处理方式的实验结果类似，在 50×50 的数据集上，当 T_0/T 较小时，SCM 和 HR-EN 表现较差；当 T_0/T 接近 1 时，MC-NNM 和 MC-WNNM 算法都表现良好。在 200×200 的数据集上，当 T_0/T 较小时，SCM 表现较差；当 T_0/T 接近 1 时，MC-NNM 和 MC-WNNM 算法表现良好。

表 3.2 不同算法在随机交错处理方式下 RMSE

	T_0	DID	SCM	VT-EN	HR-EN	MC-NNM	MC-WNNM
50×50	5	1.452202	2.014988	1.1639509	1.1511896	0.5666323	0.088666
	15	1.398726	2.145658	0.9655438	0.8280652	0.1872315	0.059508
	25	1.319963	2.229551	0.8751839	0.7273506	0.1363048	0.055741
	35	1.300635	2.280946	0.8593944	0.6126267	0.113083	0.051041
	45	1.508625	2.427556	0.8532547	0.720703	0.102555	0.050909
200×200	20	1.383386	1.674714	1.006344	1.4408218	1.428226	0.229864
	60	1.406590	1.822351	0.882504	0.8318471	0.872845	0.120063
	100	1.322830	1.790497	0.666288	0.7045295	0.297138	0.114592
	140	1.201828	1.743884	0.693443	0.4922273	0.133165	0.112191
	180	1.360767	1.710671	0.736544	0.590400	0.120337	0.112746

随机交错处理方式的误差棒图如下图 3.2 所示。

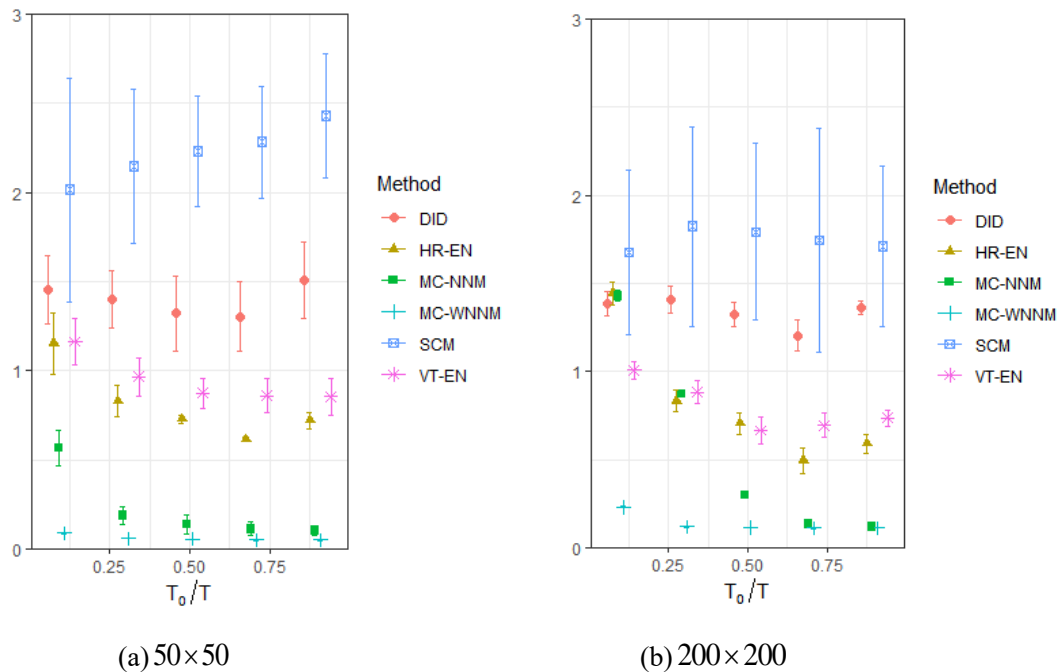


图 3.2 不同算法在随机交错处理方式下误差棒图

从图 3.2 可以看出，在随机交错处理方式下 SCM 的估计精度同样较差，MC-

NNM 和 MC-WNNM 的估计精度较高。从图 3.2 中可以明显看出, MC-WNNM 算法在不同观测率和不同规模数据集上, MC-WNNM 算法估计精度均优于其他算法。

3.5 反事实结果估计

3.5.1 数据来源

选取 Abadie、Athey 等研究中使用的有关加利福尼亚州(后文简称加州)限制吸烟政策的数据集中的控制组数据。在原数据集中, 由于加州在 1988 年实施限制吸烟政策, 其未受政策处理的值不可用, 因此将加州删去, 则 $N = 38$, $T = 31$ 。然后, 人为的指定要接受处理的个体和时间段, 并将预测值与实际值进行比较。

3.5.2 反事实结果分析

通过随机同时和随机交错处理缺失机制对数据进行矩阵填充, 并对不同观测率下的均方根误差进行对比。

表 3.3 不同算法在不同处理方式下 RMSE

	T_0	DID	SCM	VT-EN	HR-EN	MC-NNM	MC-WNNM
随机同时处理	4	29.82269	26.62686	18.59947	34.69414	19.44851	15.77530
	10	28.77200	18.38029	16.67830	26.87803	17.10940	14.73945
	16	26.55062	12.39791	12.13057	19.46361	14.60332	12.30803
	22	21.40214	13.45419	16.40956	16.46509	12.10595	11.86473
	28	16.21724	12.68395	10.58236	12.69787	10.88689	10.46804
随机交错处理	4	23.12713	28.03964	23.81269	28.57632	22.50424	16.13591
	10	22.05742	24.1544	20.49860	24.73148	16.58668	14.16867
	16	20.51406	21.60619	21.35906	19.20307	13.91697	13.38038
	22	18.43952	22.33838	17.89858	15.44863	12.35547	12.27729
	28	16.71844	21.42582	19.84682	12.48130	10.07261	9.52801

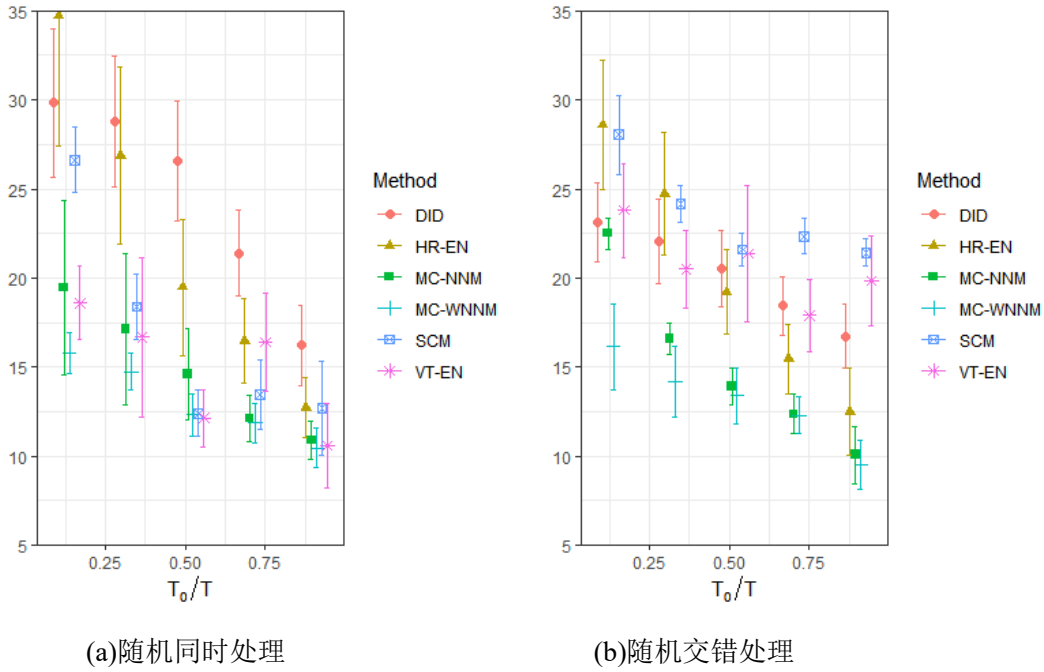


图 3.3 不同随机处理方式下误差棒图

如表 3.3 和图 3.3 所示，在随机同时处理下，当 T_0/T 较小时，HR-EN 得出的平均 RMSE 较大，而 MC-WNNM 的平均 RMSE 最小；当 T_0/T 接近 1 时，HR-EN 和 MC-WNNM 算法都表现良好。在随机交错处理情况下，当 T_0/T 较小时，HR-EN 表现较差；当 T_0/T 接近 1 时，MC-NNM 和 MC-WNNM 算法都表现良好。通过比较可以认为 MC-WNNM 算法在估计潜在结果时具有一定的稳健性。

3.6 小结

针对因果效应推断存在潜在结果缺失问题，提出使用机器学习算法，区别于传统计量经济模型，融合加权核范数正则化矩阵填充方法和因果推断方法求解约束优化问题估计缺失的潜在结果，以此提高估计的精度。利用核范数正则化思想，采用 ADMM 框架优化问题，提出加权核范数正则化矩阵填充算法。在不同规模和不同观测率的模拟数据集上，随机同时处理和随机交错处理下 MC-WNNM 均得到更精确的解，在真实数据集上得出同样结论。因此，提出的 MC-WNNM 算法是一个较好的估计潜在结果的矩阵填充算法。

4 基于倾向得分加权矩阵填充的因果效应分析

4.1 引言

在随机对照试验中大多数研究采用二元处理,当以协变量为条件时,因果效应总是异质的。因此,通常假设不存在未观测的混杂因素,而选择适当的混杂因素进行控制对因果推断至关重要。在观察性研究中,分配机制无法满足随机化条件,则样本组间均衡难以实现,因此可能导致因果效应的估计产生较大偏差。而倾向得分可以“降维”,削弱混杂因素的影响,可有效提高模型估计的效果。

倾向得分是在给定协变量的条件下,观测个体进入处理组的概率,可以将众多协变量表示为一维的向量,起到“降维”的作用。倾向得分可以最大限度地概括协变量信息,用其进行匹配可以平衡处理组和控制组在协变量上被观测到的差异。在以往的研究中,发现倾向得分模型的轻微误设,可能导致处理效应估计的显著偏差。传统方法基于线性假定计算倾向得分,较为单一。Woo 等(2010)提出当协变量间不满足线性关系时通过广义可加模型估计倾向得分效果更好。Imai(2014)提出用广义矩估计模型使处理组和控制组中协变量任意可测函数的一阶矩和二阶矩相等,得出倾向得分具有协变量平衡的优点。

随着大数据领域逐渐发展,众多研究将机器学习与倾向得分匹配方法相结合从而提高倾向得分的准确性。目前,神经网络和随机森林应用最广泛。Lee 等(2010)提到面对复杂数据随机森林对于修正选择偏差更有效,计算得分更准确。勾忠平等(2020)发现当处理因素与协变量间存在复杂关系时,基于神经网络估的倾向得分法估计平均处理效应更精确。刘展等(2021)基于随机森林倾向得分模型对网络调查数据的倾向得分进行估计,研究结果表明提出方法的估计效果比 logistic 模型好。Athey 等(2021)提出基于矩阵填充算法计算倾向得分并加权损失函数以平衡样本。Poulos(2023)在 Athey 工作基础上实现矩阵填充估计倾向得分方法,利用倾向得分对每个个体赋予相应的权重进行加权,构造平衡样本消除混杂因素的影响。

本章在 Athey 及 Poulos 工作基础上,结合机器学习中随机森林分类和降维的特点,提出基于随机森林倾向得分模型加权调整矩阵填充损失函数,解决两种潜在结果的协变量分布不平衡问题。

4.2 方法介绍

4.2.1 倾向得分

在观测性研究数据中，为了不遗漏混杂变量，需要将可能的协变量都纳入模型内，这样虽然能使所有混杂变量都被观测到，但也可能会引入过多干扰信息。倾向得分方法是建立在反事实框架下的一种有效平衡数据的方法。因此，基于倾向得分的估计量越来越多地应用于因果推断。表示在给定协变量的条件下，定义倾向得分为 $e_{ii} = \Pr(W_{ii} = 1 | X_{ii})$ ，表示观测个体进入处理组的概率。因此，倾向得分包含了协变量信息，可以通过其平衡处理组和控制组间被观测到的差异。

基于倾向得分的方法可以分为匹配和加权两类，基于倾向得分的匹配方法使用倾向得分来构建匹配集合，将倾向得分相似的个体放到一个集合中，每个集合中同时包含处理组个体和控制组个体，并且集合中个体具有相似分布；基于倾向得分的加权方法使用倾向得分来创造权重，对观测协变量进行平衡。逆概率加权法是一种基于倾向得分的加权方法，将倾向得分的倒数作为个体权重，单个个体 i 的权重记为：

$$w_i = \frac{W_i}{e(x_i)} + \frac{1 - W_i}{1 - e(x_i)},$$

即处理组个体权重为 $w_i = \frac{W_i}{e(x_i)}$ ，控制组个体权重为 $w_i = \frac{1 - W_i}{1 - e(x_i)}$ 。使用逆概率

加权方法估计 ATE：

$$ATE = E \left[\frac{W_i Y_i}{e(X_i)} - \frac{(1 - W_i) Y_i}{1 - e(X_i)} \right]$$

逆概率加权方法通过加权方法平衡处理组和对照组，使观察性研究数据尽可能接近随机对照试验数据，从而减少使用观测性研究数据进行因果推断时产生的偏倚。

基于倾向得分的方法对修正样本选择偏差一般要满足以下假设：

假设 4(强可忽略): $[Y(0), Y(1)] \perp\!\!\!\perp W | X$.

强可忽略假设要求当以协变量为条件时潜在结果与处理分配机制无关，并意

意味着不存在无法测量的混杂因素。根据假设 4 的含义以及倾向得分的定义可得： $[Y(0), Y(1)] \perp\!\!\!\perp W | e(X)$ ，其中 $e(X) = \Pr(W = 1 | X)$ 表示倾向得分。

假设 5(共同区间): 对所有协变量 X 有 $0 < \Pr\{W = 1 | X\} < 1$ 。

共同区间假设要求个体分配到处理组的概率限制在 0 到 1 之间，且要求处理组和控制组的协变量分布之间有重叠。

4.2.2 随机森林倾向得分模型

传统的倾向得分估计方法依赖于模型的正确设定，而利用随机森林模型估计倾向得分能有效避免传统模型设定匹配结果的平衡性。随机森林是基于决策树分类的一种集成学习方法，相比于其他算法被证明预测精度高，泛化能力强。随机森林的重要特点是通过随机选择样本和随机选择特征构建子集。

假设观测数据集 $U = \{Y_i, W_i, X_i\}$ ，其中 Y_i 为观测结果变量， W_i 为二分处理变量 ($W_i = 1$ 表示处理组， $W_i = 0$ 表示控制组)， X_i 为协变量，则在给定 X_i 条件下样本进入处理组的概率 $p_i = P(W_i = 1 | X_i)$ ，本质上就是倾向得分。随机森林估计倾向得分包括以下四个步骤：

(1) 以 Bagging 为基础，从数据集 U 中有放回地随机抽取 N_{tree} 个子样本作为新的训练集，从特征 X_i 中有放回地随机抽取 N_{tree} 个协变量集；

(2) 根据训练集和对应的协变量集生成 N_{tree} 个 CART 决策树，组成随机森林；

(3) 根据各单棵树的分类结果计算样本被分处理组的概率 p_i ；

(4) 将从 N_{tree} 棵决策树中得到的概率值 p_i 进行平均，获得每个样本的随机森林倾向得分 $p(X_i)$ 。

通过随机森林估计到倾向得分是决策树中每层协变量的选择 $W_i = 1$ 的概率，与传统倾向得分定义相同。

在观察性研究中，处理效应表示为： $\tau_{ATT} = E(Y_1 | W_i = 1) - E(Y_0 | W_i = 1)$ ，其

中反事实结果 Y_0 无法直接观测到, 大多数研究中利用控制组的数据估计 Y_0 。当协变量存在异质性时, 使用倾向得分平衡处理组和控制组混杂变量分布。假设控制组和处理组混杂变量分布服从 $f(X_i | W_i = 1)$ 和 $f(X_i | W_i = 0)$, 设权重函数 $W(\mathbf{X})$, 使得两组之间混杂变量分布达到平衡

$$f(X_i | W_i = 1) = W(\mathbf{X}) f(X_i | W_i = 0), \quad (4.1)$$

其中 $W(\mathbf{X}) = \frac{p(X_i)}{1 - p(X_i)}$, 即处理组个体权重为 1, 控制组个体权重为 $\frac{p(X_i)}{1 - p(X_i)}$ 。

4.2.3 倾向得分加权损失函数矩阵填充模型

使用矩阵填充的目的是利用控制组数据估计处理组个体的潜在结果, 将控制组结果定义为:

$$Y_{it} = L_{it} + \gamma_i + \delta_t + \varepsilon_{it}, \quad (4.2)$$

其中, L_{it} 表示含缺失的观测到的结果; γ_i 表示个体固定效应; δ_t 表示时间固定效应; ε_{it} 表示残差项, 满足严格外生性假设。Athey 等(2021)提出的矩阵填充模型:

$$\hat{\mathbf{L}} = \arg \min_{L, \gamma, \delta} \frac{1}{|O|} \sum_{(i,t) \in O} (Y_{it} - L_{it} - \gamma_i - \delta_t)^2 + \lambda_L \|\mathbf{L}\|_* \quad (4.3)$$

其中, $\|\cdot\|_* = \sum_i \sigma_i(\cdot)$ 表示核范数, 为奇异值之和; λ_L 为调节参数, 由 K 折交叉验证得到; $(i,t) \in O$ 表示观测到的位置索引。Poulos(2023)通过 lasso 逻辑回归计算倾向得分 \hat{w}_{it} , 对模型(4.3)中损失函数进行加权:

$$\hat{\mathbf{L}} = \arg \min_{L, \gamma, \delta} \frac{1}{|O|} \sum_{(i,t) \in O} \frac{\hat{w}_{it}}{1 - \hat{w}_{it}} (Y_{it} - L_{it} - \gamma_i - \delta_t)^2 + \lambda_L \|\mathbf{L}\|_* \quad (4.4)$$

基于模型(4.4), 本节提出利用随机森林估计倾向得分, 再对加权核范数矩阵填充的损失函数进行加权调整:

$$\hat{\mathbf{L}} = \arg \min_{L, \gamma, \delta} \frac{1}{|O|} \sum_{(i,t) \in O} \frac{\hat{e}_{it}}{1 - \hat{e}_{it}} (Y_{it} - L_{it} - \gamma_i - \delta_t)^2 + \lambda_L \|\mathbf{L}\|_{w,*} \quad (4.5)$$

其中, \hat{e}_{it} 为随机森林倾向得分。

4.3 数值模拟

4.3.1 数据生成

在随机森林估计倾向得分方法中，模拟样本量为 $N = 200$ 的数据集，假定协变量 \mathbf{X} 服从均值为 0，方差为 1，协方差为 0.1 的多元正态分布，即

$$\mathbf{X} = (X_1, X_2, X_3, X_4, X_5)^\top \sim N_5(\mu, \Sigma),$$

其中

$$\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 1 & 0.1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 1 & 0.1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 1 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.1 & 1 \end{pmatrix}.$$

处理变量与协变量间可能存在不同程度潜在线性、非线性和交互作用等关系，在模拟中指定处理变量 \mathbf{D} 和协变量 \mathbf{X} 之间为线性关系，假设处理变量 D_i 服从多维多元线性模型：

$$D_i = \beta^\top X_i + \varepsilon_i, \quad i = 1, 2, \dots, N,$$

则假定二元处理变量 \mathbf{D} 为：

$$D_i = X_{i1} + X_{i2} + 0.5X_{i3} + 0.5X_{i4} + X_{i5}.$$

一般采用线性模型生成连续的结果变量 \mathbf{Y} ：

$$Y_i = D_i + X_{i1} + X_{i2} + 0.5X_{i3} + 0.5X_{i4} + 0.3X_{i5} + \varepsilon_i,$$

其中 $\varepsilon_i \sim N(0, 2^2)$ 。

4.3.2 随机森林估计倾向得分

根据模拟数据集利用随机森林估计倾向得分，利用模拟协变量 \mathbf{X} 进行估计，根据处理变量 \mathbf{D} 与 5 个协变量的关系，计算样本被分处理组的概率。随机森林模型可以用 ROC 曲线来评价模型的性能，如图 4.1 所示。

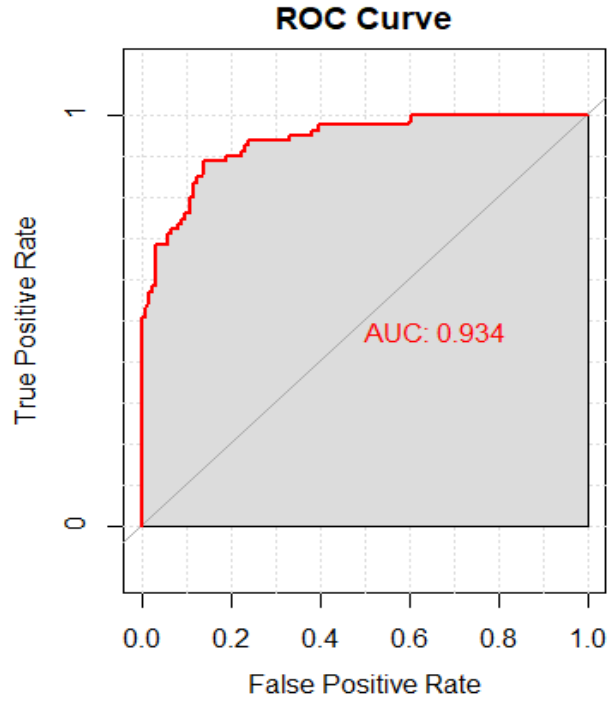


图 4.1 随机森林模型 ROC 曲线图

由图 4.1 可知，ROC 曲线很靠近左上角，表示随机森林模型分类的准确性很高。同时，ROC 曲线下的面积 $AUC=0.934$ ，也充分说明随机森林模型预测分类效果好。随机森林模型在使用袋外样本进行交叉验证时，模型对于未在训练集中采样的数据分类错误率为 1.55%，换言之分类的准确率为 98.45%。

随机森林估计倾向得分后需要检验协变量的平衡性。对于连续型变量通常采用 t 检验比较两组协变量分布是否相同。

表 4.1 协变量平衡情况

	加权前		RF 加权	
	t 检验	p 值	t 检验	p 值
X_1	-8.6338	$3.752e-15^{***}$	-1.5784	0.1217
X_2	-10.222	$2.2e-16^{***}$	-1.4055	0.1665
X_3	-9.0044	$4.953e-16^{***}$	-1.5431	0.1290
X_4	-8.2189	$5.141e-14^{***}$	0.78933	0.4340
X_5	-11.143	$2.2e-16^{***}$	-0.80106	0.4267

注：***、**、*分别代表 1%、5%、10%的显著性水平。

在进行假设检验时，若 p 值小于预设的显著性水平 $\alpha = 0.01$ ，则可拒绝原假设，认为两组协变量有明显差异，即没有达到平衡。根据上表 4.1 所示，从 p 值大小层面判断，倾向得分加权前这 5 个协变量的 p 值均小于 0.01，说明协变量的分布存在显著差异，即可以认为原始样本协变量不平衡。而采用随机森林模型估计的倾向得分加权后，协变量 p 值均大于 0.01，均通过 t 检验，即接受原假设，可认为倾向得分加权后的协变量满足平衡性条件。

4.3.3 因果效应分析

经过上述操作随机森林估计的倾向得分可以平衡协变量分布，下一步提取随机森林模型估计的倾向得分 \hat{e}_{it} ，根据式(4.1)计算矩阵填充损失函数权重

$$W_{it}(X) = \frac{\hat{e}_{it}}{1 - \hat{e}_{it}}.$$

在模拟数据集上，估计缺失数据的反事实结果并估计平均处理效应，结果如表 4.2 所示。

表 4.2 不同方法的 RMSE 及平均处理效应

方法	RMSE	ATT
DID	8.386243	5.816843
SCM	9.335655	6.559987
MCNNM	5.189882	4.149634
EN	9.814562	6.813374
ENT	11.538124	7.631384
RF-WNNM	5.168553	4.143242

从表 4.2 结果可以看出，处理变量对结果变量的影响均是正向的，水平和垂直回归的弹性网络方法得出的结果误差较大，而矩阵填充方法预测结果的误差较小，其中提出的随机森林倾向得分加权损失函数方法的 RMSE 最小，平均处理效应为 4.143242。总体来看，矩阵填充方法预测反事实结果的误差更小。随着不同方法得到的误差越小，而平均处理效应也越小。

4.4 基于倾向得分加权的因果效应分析

4.4.1 数据来源与处理

Abadie 等(2010)研究了美国加州香烟控制 99 号法案对加州香烟消费的抑制作用。加州为处理组,美国其他未出台相似政策的州为控制组。收集了包含加州在内 39 个州从 1970 年到 2000 年的州级层面年度面板数据。选取的变量说明如下:

表 4.3 变量说明

	变量名称	变量类型	变量含义
结果变量	cigsale	连续变量	州年度人均香烟消费量
处理变量	status	0-1 变量	0 表示未受干预, 1 表示接受干预
	retprice	连续变量	香烟平均零售价格
	lnincome	连续变量	州人均收入对数
	age15to24	连续变量	州人口 15-24 岁人口占比
协变量	beer	连续变量	州人均啤酒消费量
	cig75	连续变量	1975 年人均香烟消费量
	cig80	连续变量	1980 年人均香烟消费量
	cig88	连续变量	1988 年人均香烟消费量

在后续研究中计算倾向得分要求样本无缺失值,但该数据存在协变量缺失情况,因此需要对缺失值进行处理。在 Abadie 等(2010)的研究中采用均值插补,观察数据缺失特征,发现为条块缺失,基于协变量间的相关性本节采用时间序列线性插补方法填补缺失值。

4.4.2 倾向得分平衡协变量

本研究将加州禁烟数据采用随机森林算法估计倾向得分。倾向得分的关键是利用已知的协变量进行估计,因此根据已知插补后的协变量输入随机森林模型,根据处理变量与 7 个协变量的关系,计算样本被分处理组的概率。随机森林模型可以用 ROC 曲线来评价模型的性能,如图 4.1 所示。

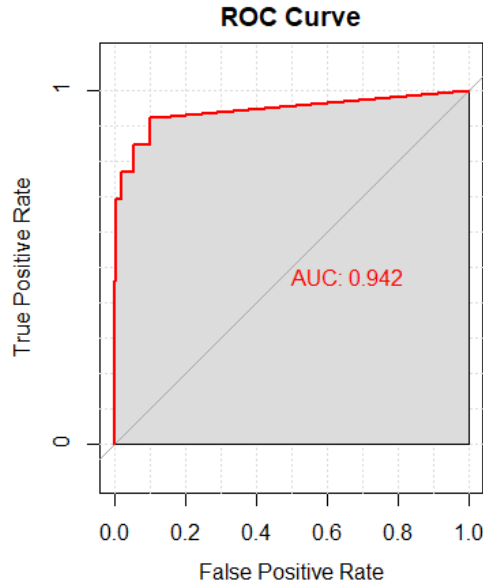


图 4.2 随机森林模型 ROC 曲线图

由图 4.2 可知，ROC 曲线非常靠近左上角，说明随机森林模型分类的准确性很高。同时，ROC 曲线下的面积 $AUC=0.942$ ，也充分说明随机森林模型预测分类效果好。随机森林模型在使用袋外样本进行交叉验证时，模型对于未在训练集中采样的数据分类错误率为 1.08%，换言之分类的准确率为 98.82%。

表 4.4 倾向得分加权的协变量平衡性检验

	加权前		RF 加权	
	t 检验	p 值	t 检验	p 值
lnincome	-11.599	2.792e-08***	-0.047252	0.9629
beer	3.8556	0.001968***	-1.6266	0.1221
age15to24	5.7672	7.903e-05***	1.4661	0.1559
retprice	-5.7911	7.688e-05***	-1.4206	0.1683
cig75	9.0841	2.2e-16***	0.26703	0.794
cig80	20.644	2.2e-16***	0.34597	0.7354
cig88	33.667	2.2e-16***	1.1032	0.2916

注：***、**、*分别代表 1%、5%、10%的显著性水平。

根据上表 4.4 所示，从 p 值大小层面判断，倾向得分加权前这 7 个协变量的 p 值均小于 0.01，说明协变量的分布存在显著差异，即可以认为原始样本协变量不平衡。而采用随机森林模型估计的倾向得分加权后，协变量 p 值均大于 0.01，

均通过 t 检验，即接受原假设，在统计学意义上可认为倾向得分加权后协变量满足平衡性条件。

4.4.3 平均处理效应估计

经随机森林倾向得分加权使得协变量具有可比性。提取上节估计出的倾向得分 \hat{e}_{it} ，根据式(4.1)计算权重

$$W_{it}(X) = \frac{\hat{e}_{it}}{1 - \hat{e}_{it}}.$$

在加州禁烟数据集上，为研究加州在 1989 年正式实施控烟政策效果，Abadie 等提出的合成控制法的重要思想是通过协变量合成与加州相似的控制组，即加州的反事实估计。下面利用模型(4.4)矩阵填充算法估计加州的反事实结果并估计平均处理效应，结果如表 4.5 所示。

表 4.5 不同方法的 RMSE 及平均处理效应

方法	RMSE	ATT
DID	27.99652	-26.48595
SCM	19.78401	-18.41649
MCNNM	21.32916	-19.51129
EN	12.85314	11.71136
ENT	20.40224	17.75024
RF-WNNM	12.50763	-5.220282

从表 4.5 结果可以看出，水平和垂直回归的弹性网络方法结果显示平均处理效应为正向，而已知文献中结果显示平均处理效应为负向，即控烟政策能抑制香烟销量。本章提出的随机森林倾向得分加权损失函数方法的误差 RMSE 最小，平均处理效应为-5.220282，相较于传统方法预测效果更好。

4.5 小结

本章通过已有文献，确定因果推断的最终目的是根据已观测数据估计反事实结果，提出采用随机森林模型估计倾向得分，再根据倾向得分协变量平衡性质对矩阵填充模型损失函数进行加权，并估计平均处理效应。为了验证随机森林倾向

得分矩阵填充模型的有效性,基于加州控烟真实数据集将提出的方法与双重差分法、合成控制法、核范数矩阵填充模型、水平和垂直弹性网络模型估计反事实,研究结果表明随机森林倾向得分矩阵填充模型同样适用于反事实预测,且相比较其预测误差更小。

5 总结与展望

5.1 总结

随着大数据科学的发展,越来越多机器学习方法引入到传统计量经济学方法中。在潜在结果框架下,针对因果推断中潜在结果缺失问题,研究正则化矩阵填充算法预测反事实结果并估计因果推断关注的平均处理效应。主要工作包括两部分:

一是在不考虑协变量的情况下,针对已有的核范数正则化矩阵填充算法估计反事实结果时对所有奇异值用同一值进行收缩问题,提出使用加权核范数矩阵填充算法,利用自适应加权奇异值的方法充分保留数据的有效信息;进一步利用ADMM 框架优化问题,使其具有理论依据。同时,采用因果推断中常用的随机同时和随机交错处理方式在模拟数据集上验证加权核范数估计反事实结果的可行性,数据模拟显示,基于加权核范数的矩阵填充算法在不同规模和观测率的模拟数据集上可得到更精确的解。加权核范数的矩阵填充算法同样适用于真实数据集,发现与传统计量方法相比,该方法能更精确预测反事实结果。

二是考虑协变量对结果变量的影响。在观察性研究中,通常处理组和控制组的协变量分布不平衡,而倾向得分具有协变量平衡性质,因此可利用倾向得分解决矩阵填充模型对于观测结果变量与预测变量的差值赋予同等权重问题。然而,传统的倾向得分估计方法依赖模型的正确选择,随机森林算法能很好的解决这一缺陷。因此,提出采用随机森林模型估计倾向得分,再根据倾向得分协变量平衡性质对矩阵填充模型损失函数进行加权,并估计平均处理效应。在模拟数据集和真实数据集上验证随机森林倾向得分矩阵填充模型的有效性,发现提出的方法同样适用于反事实预测,且相比较其预测误差更小。

5.2 展望

关于因果推断,从潜在结果缺失的角度出发,对矩阵填充模型进行一些改进工作,但将机器学习算法融合进因果推断中仍处于理论分析阶段,在实际应用中仍存在较多困难。

(1)提出的加权核范数矩阵填充模型依赖于奇异值分解,当恢复高维数据矩

阵时，进行奇异值分解的成本巨大。在后续研究中，需要寻找更适用的矩阵分解方法。

(2) 目前，针对协变量平衡的倾向得分法可应用于高维数据，但随机森林估计方法应用不同数据集的差异需要进一步分析。并且在实际应用中，面临着协变量数据缺失问题，也是后续研究需要解决的地方。

参考文献

- [1] Abadie A , Gardeazabal J .The Economic Costs of Conflict: A Case Study of the Basque Country[J].American Economic Review, 2003, 93(1):113-132.
- [2] Abadie A , Diamond A , Hainmueller J .Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program[J].Journal of the American Statistical Association, 2010, 105(490):493-505.
- [3] Abadie A .Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects[J].Journal of Economic Literature, 2021, 59(2):391-425.
- [4] Abadie A , L'Hour J.A Penalized Synthetic Control Estimator for Disaggregated Data[J].Journal of the American Statistical Association, 2021, 116(536):1817-1834.
- [5] Amjad M J , Shah D , Shen D .Robust Synthetic Control[J]. Journal of Machine Learning Research,2018,19(1):802-852.
- [6] Athey S, Bayati M, Imbens G, et al. Ensemble methods for causal effects in panel data settings[C]//AEA papers and proceedings. 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association, 2019, 109: 65-70.
- [7] Athey S , Bayati M , Doudchenko N ,et al.Matrix Completion Methods for Causal Panel Data Models[J].Journal of the American Statistical Association, 2021, 116(536):1716-1730.
- [8] Bai J .Panel Data Models With Interactive Fixed Effects[J].Econometrica, 2009, 77(4):1229-1279.
- [9] Bai J , Ng S .Matrix Completion, Counterfactuals, and Factor Analysis of Missing Data[J].Journal of the American Statistical Association, 2021, 116(536):1746-1763.
- [10] Ben-Michael Eli, Avi Feller, Jesse Rothstein. The Augmented Synthetic Control Method[J]. Journal of the American Statistical Association,2021,1789-1803.
- [11] Ben-Michael E , Feller A , Rothstein J .Synthetic Controls with Staggered Adoption[J].NBER Working Papers, 2021.
- [12] Cai J , Candès E J , Shen Z .A Singular Value Thresholding Algorithm for Matrix

- Completion[J].SIAM Journal on Optimization, 2010,20(4):1956-1982.
- [13] Cai J F , Osher S .Fast Singular Value Thresholding without Singular Value Decomposition[J].Methods and applications of analysis, 2013,20(4):335-352.
- [14] Callaway B , Sant'Anna P H C .Difference-in-Differences with multiple time periods[J].Journal of Econometrics, 2021,225(2):200-230.
- [15] Candès E J,Recht B.Exact matrix completion via convex optimization[J].Foundations of Computational Mathematics, 2009,9(6):717-772.
- [16] Chaisemartin C D , D'Haultfoeuille X .Two-way Fixed Effects Regressions with Several Treatments[J/OL].2020:arXiv:2012.10077.
- [17] Cheng J , Wyss R , Franklin J M ,et al.Collaborative-controlled LASSO for constructing propensity score-based estimators in high-dimensional data[J].Sage Publications, 2019(4):1-34.
- [18] Fan J, Imai K, Liu H, et al. Optimal Covariate Balancing Conditions in Propensity Score Estimation[J]. Journal of Business & Economic Statistics,2022, 41(1): 97–110.
- [19] Fazel M , Hindi H , Boyd S P .Log-det heuristic for matrix rank minimization with applications to Hankel and Euclidean distance matrices[C]//Proceedings of the 2003 American Control Conference. Denver: IEEE,2003:2156-2162.
- [20] Gamarnik D, Misra S. A note on alternating minimization algorithm for the matrix completion problem[J]. IEEE Signal Processing Letters, 2016, 23(10): 1340-1343.
- [21] Gu S, Zhang L, Zuo W, et al. Weighted nuclear norm minimization with application to image denoising[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 2862-2869.
- [22] Gu S H, Xie Q, Meng D Y, et al. Weighted nuclear norm minimization and its applications to low level vision[J]. International Journal of Computer Vision, 2017, 121(2): 183-208.
- [23] Hu Y , Zhang D , Liu J ,et al.Accelerated singular value thresholding for matrix completion[C]//Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: Association for Computing Machinery, 2012:298-306.

- [24] Imai K, Ratkovic M. Covariate Balancing Propensity Score[J]. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 2014,76(1): 243–263.
- [25] Jin S , Miao K , Su L.On factor models with random missing: EM estimation, inference, and cross validation[J].*Journal of Econometrics*, 2021, 222(1):745-777.
- [26] Josey K, Juarez-Colunga E, Yang F, et al. A Framework for Covariate Balance Using Bregman Distances[J]. *Scandinavian Journal of Statistics*,2021, 48(3): 790–816.
- [27] Lee B K , Lessler J , Stuart E A .Improving propensity score weighting using machine learning[J].*Statistics in Medicine*, 2010, 29(3):337-346.
- [28] Li F,Morgan K, Zaslavsky A.Balancing Covariates via Propensity Score Weighting[J].*Journal of the American Statistical Association*,2018,113(521): 390–400.
- [29] Li F, Li F. Propensity Score Weighting for Causal Inference with Multiple Treatments[J]. *Annals of Applied Statistics*, 2019, 13(4): 2389–2415.
- [30] Mao X , Chen S X , Wong R K W .Matrix Completion With Covariate Information[J].*Taylor & Francis*, 2019,114(525):198-210.
- [31] Mazumder R , Hastie T , Tibshirani R .Spectral Regularization Algorithms for Learning Large Incomplete Matrices[J]. *Journal of Machine Learning Research*, 2010,11(80):2287-2322.
- [32] Ning Y, Peng S, Imai K. Robust Estimation of Causal Effects via a High-dimensional Covariate Balancing Propensity Score[J]. *Biometrika*, 2020,107(3): 533–554.
- [33] Pesaran M H .Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure[J].*Econometrica*, 2006, 74(4):967-1012.
- [34] Poulos J .State-Building through Public Land Disposal? An Application of Matrix Completion for Counterfactual Prediction[J].*Papers*, 2019.DOI:10.48550/arXiv.1903.08028.
- [35] Rosenbaum P R, Rubin D B. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score[J]. *The American Statistician*, 1985, 39(1): 33-38.

- [36] Schnabel T , Swaminathan A , Singh A ,et al.Recommendations as Treatments: Debiasing Learning and Evaluation[C]// Proceedings of the 33rd International Conference on International Conference on Machine Learning, PMLR ,2016,48:1670-1679.
- [37] Shen Z , Cui P , Kuang K ,et al.Causally Regularized Learning with Agnostic Data Selection Bias[C]// Proceedings of the 26th ACM international conference on Multimedia ,2018,1-9.
- [38] Toh K C , Yun S .An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems[J].Pacific Journal of Optimization, 2010, 6(3):615-640.
- [39] Wang J , Wong R K W , Mao X ,et al.Matrix Completion with Model-free Weighting[C]// Proceedings of the 38th International Conference on Machine Learning, PMLR , 2021,139:10927-10936.
- [40] Wang X , Zhang R, Sun Y ,et al. Doubly Robust Joint Learning for Recommendation on Data Missing Not at Random [C]// Proceedings of the 36th International Conference on Machine Learning, PMLR , 2019,97:6638-6647.
- [41] Woo M J , Reiter J P , Karr A F .Estimation of propensity scores using generalized additive models[J].Statistics in Medicine, 2010, 27(19):3805-3816.
- [42] Xiong R, Pelger M. Large dimensional latent factor modeling with missing observations and applications to causal inference[J]. Journal of Econometrics, 2023, 233(1): 271-301.
- [43] 丁杨. 因果推断在科学计量学中的应用研究[D].山东师范大学,2021.
- [44] 董珍. 多重政策因果效应的评估方法及应用研究[D].天津财经大学,2019.
- [45] 范菊逸,詹铭峰,蔡宗武等.带有变量选择的协变量平衡倾向得分的估计:基于 GMM-LASSO 方法[J].系统工程理论与实践,2021,41(10):2631-2639.
- [46] 冯雅莉,孙为军.一种基于快速随机投影的矩阵填充方法[J].计算机应用与软件,2019,36(09):106-110+121.
- [47] 高倩,王彤.高维情况下基于倾向性评分的因果推断方法[J].中国卫生统计,2021,38(06):945-949.

- [48] 勾忠平,向良成,梁秀芳等.神经网络在倾向分数值估计中的研究[J].中国循证医学杂志,2020,20(09):1075-1082.
- [49] 韩锋.协变量存在缺失的因果效应稳健估计[J].统计与决策,2020,36(14):37-39.
- [50] 蒋青嬿,马佳羽,黄灿等.因果推断中基于能源距离的协变量分布平衡[J].统计研究,2023,40(05):144-151.
- [51] 邝神芬,黄业文,宋杰等.基于深度矩阵分解网络的矩阵填充方法[J].计算机科学,2019,46(10):55-62.
- [52] 李超,求文星.基于机器学习的因果推断方法研究进展[J].统计与决策,2021,37(11):10-15.
- [53] 刘展,金勇进.基于倾向得分匹配与加权调整的非概率抽样统计推断方法研究[J].统计与决策,2016,No.465(21):4-8.
- [54] 刘展,潘莹丽,金美玲.大数据背景下网络调查样本的随机森林倾向得分模型推断研究[J].统计研究,2021,38(11):130-140.
- [55] 欧祥威,董春雨.从大数据揭示因果机制的可能性[J].系统科学学报,2022,30(04):26-31.
- [56] 潘伟,胡春安.基于低秩矩阵填充的推荐算法[J].科学技术与工程,2021,21(11):4519-4523.
- [57] 潘伟,胡春安.基于加权 Schatten-p 范数的矩阵填充及其应用[J].计算机应用与软件,2023,40(04):230-235.
- [58] 彭非,吴浩.基于倾向性得分匹配法的平均处理效应的自助法推断[J].统计与信息论坛,2019,34(08):12-19.
- [59] 钱浩祺,龚嫣然,吴力波.更精确的因果效应识别:基于机器学习的视角[J].计量经济学报,2021,1(04):867-891.
- [60] 沈艳,李星宇,周前坤.大数据背景下面板数据政策评估的估计和推断[J].数量经济技术经济研究,2022,39(06):120-139.
- [61] 石莹,黄华,王智等.基于加权核范数的低秩矩阵补全算法研究[J].西南大学学报(自然科学版),2022,44(05):192-202.
- [62] 孙乐.基于矩阵填充与 BP 神经网络的混合推荐算法[J].信息与电脑(理论版),2020,32(21):41-43.

- [63] 邵凌楠,王春雨,田茂再.缺失数据下的逆概率多重加权分位回归估计及其应用[J].统计研究,2018,35(09):115-128.
- [64] 王俊霞,申倩影,王川龙.符号矩阵填充的修正增广拉格朗日乘子算法[J].工程数学学报,2021,38(03):343-352.
- [65] 王志远,王兴芬.基于用户兴趣差异改进矩阵填充的个性化推荐算法[J].计算机应用与软件,2020,37(12):224-230+237.
- [66] 温瑞萍,肖云,王川龙.基于均值修正的 Toeplitz 矩阵填充的增广拉格朗日乘子算法[J].数值计算与计算机应用,2022,43(01):61-75.
- [67] 吴浩,彭非.基于协变量平衡加权的平均处理效应的稳健有效估计[J].统计研究,2020,37(04):114-128.
- [68] 肖云,温瑞萍.Toeplitz 矩阵填充的尾端修正增广拉格朗日乘子算法[J].首都师范大学学报(自然科学版),2022,43(05):8-15.
- [69] 薛娇,傅德印,韩海波等.基于多视角学习的非负函数型矩阵填充算法[J].统计与决策,2022,38(07):5-11.
- [70] 张晨,肖君儒,周丽.基于矩阵填充的混合协同过滤推荐算法研究[J].数学的实践与认识,2021,51(10):81-89.
- [71] 张玮奇,张宏志,左旺孟等.基于加权核范数最小化的矩阵填充模型[J].计算机科学,2015,42(07):254-257+290.
- [72] 张娅楠,赵建伟,曹飞龙.加权核范数的矩阵恢复正则化算法[J].中国计量大学学报,2016,27(04):471-479.

攻读硕士学位期间承担的科研任务及主要成果

已发表论文:

[1]牛成英,邢晓文,闫新宇.基于 Mask R-CNN 和样方密度法的城市功能区识别[J].湖北民族大学学报(自然科学版),2023,41(03):405-413.

参与科研项目:

甘肃省中央引导地方科技发展资金项目《城市计算方法体系构建及甘肃智慧城市应用》(项目编号: YDZX20216200001876)。

致谢

花开花落万物道，聚散离别终有时。行文至此，三年的研究生学习阶段即将结束，回首过往，有万般不舍，也有许多要感谢的人。

首先要由衷地感谢我的导师牛成英教授。在我攻读硕士学位期间，牛成英教授给予我无私的指导和悉心的关怀，她在学术研究上的丰富知识和严谨态度，对我产生了巨大的影响。在整个论文写作过程中，牛成英教授给予我充分的自由度和鼓励，帮助我培养创新思维和独立解决问题的能力，而且多次细心的帮我修改，使我受益良多。

感谢我的家人和朋友们。很幸运我拥有一个幸福的家庭，感谢我的爸爸妈妈给予我的无限包容，感谢我的姐姐在我人生道路上的指引。感谢一起学习、奋斗的同门和室友，感谢你们一直以来的支持和鼓励。

再次感谢所有在我学业道路上给予我帮助和支持的人。我们虽不期而遇，但未来必定可期。