DOI:10.11918/202411088

多元处理情景下的异质性处理效应估计方法

马大明[1],王佳宁[2],王 楠[2]

(1.中国大连高级经理学院 教学科研部,辽宁 大连 116024;2.大连理工大学 经济管理学院,辽宁 大连 116024)

摘 要:为解决现有因果异质性效应估计方法主要聚焦于二元处理而难以适应多元处理情景的不足,提出一种创新的两阶段多元处理-因果森林(multiple treatment-causal forest,MT-CF)方法。该方法通过两个阶段的分析过程实现对多元处理因果异质性效应的精准估计。第1阶段,采用传统因果森林方法计算成对处理之间的效应差异;第2阶段,将成对处理的效应差异拓展至多元处理,从而估计每种处理相较于所有处理平均效应的差异。设计两组模拟实验以验证MT-CF方法在不同情景下的表现;将MT-CF方法应用于监测、流行病学和最终结果(the surveillance,epidemiology,and end results,SEER)数据库中,通过对乳腺癌数据的分析,评估患者术后辅助治疗的个体化疗效。结果表明:在两组模拟实验中,MT-CF方法均表现优异,其多个评价指标均优于其他基准方法;在评估患者术后辅助治疗的个体化疗效中,该方法也能估计多种治疗方案的相对效应,为患者个体化治疗决策提供科学依据,同时进一步验证了MT-CF方法在真实医疗数据中的实用性和可靠性。所提方法不仅为多元处理因果效应的估计提供一种有效工具,也为复杂因果推断问题的解决提供了新的视角与思路。

关键词:因果机器学习;因果森林;多元处理;异质性处理效应;个体化疗效评估

中图分类号:TP181 **文献标志码**:A **文章编号**:0367-6234(2026)09-0163-10

Estimation method of heterogeneous treatment effect with multiple treatment

Ma Daming[1], Wang Jianing[2], Wang Nan[2]

(1. Department of Teaching and Research, China Business Executives Academy, Dalian, Dalian 116024, Liaoning, China;
2. School of Economics and Management, Dalian University of Technology, Dalian 116024, Liaoning, China)

Abstract: To address the limitations of existing causal heterogeneity effect estimation methods, which primarily focus on a binary treatment and fail to accommodate multiple treatment scenarios, this paper proposes an innovative two-stage multiple treatment-causal forest (MT-CF) method. This method achieves precise estimation of causal heterogeneity effects in multiple treatment settings through a two-stage analytical process. In the first stage, traditional causal forest methods are used to estimate pairwise treatment effect differences. In the second stage, these pairwise differences are extended to the multiple treatment context, enabling the estimation of the difference between each treatment and the average effect across all treatments. Two sets of simulation experiments were designed to verify the performance of the MT-CF method in different scenarios. In addition to the simulation experiments, the MT-CF method was applied to the analysis of the breast cancer data in the database of the surveillance, epidemiology, and end results (SEER) to evaluate the individualized treatment effects of postoperative adjuvant therapies. The results indicate that the MT-CF method performed well in both simulation experiments, with multiple evaluation metrics outperforming other benchmark methods. In evaluating the individualized treatment effects of postoperative adjuvant therapies, the MT-CF method can also estimate the relative effects of various treatment options. This provides a scientific basis for making individualized treatment decisions. Furthermore, the results validate the MT-CF method's practicality and reliability when applied to real-world medical data. The proposed method not only provides an effective tool for estimating multiple treatment causal effect, but also provides a new perspective and idea for solving complex causal inference problems.

Keywords: causal machine learning; causal forest; multiple treatment; heterogeneity treatment effect; individualized treatment effect evaluation

异质性处理效应(heterogeneous treatment effect,HTE)的精准估计在个体化医疗、定制化营销推荐、社会政策评估等诸多领域中具有重要的应用价值。由于个体特征的差异性,不同个体或亚组对相同处

收稿日期:2024-11-28;**录用日期**:2024-12-30;**网络首发日期**:2025-05-22
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250522.1623.002
作者简介:马大明(1982—),男,副研究员
通信作者:王 楠,wnan66@mail.dlut.edu.cn

理的响应可能显著不同,一些个体或亚组可能显著受益,而另一些则可能无反应甚至受到不良影响。识别并量化这种异质性处理效应对于制定有效的个性化干预策略至关重要,有助于确保每个个体或亚组获得最佳的处理方案。

随机对照实验(randomized controlled trial,RCT)被认为是揭示处理与结果因果关系的金标准[1],但RCT面临招募困难和成本高昂等挑战,研究人员往往难以实施。随着大数据的兴起和机器学习技术的进步,观测性数据已成为替代RCT进行因果推断的可行方案,并且已有许多具有良好统计特性的因果估计方法被提出,如贝叶斯加性回归树(Bayesian additive regression trees, BART)[2]、因果树(causal tree,CT)[3]、因果森林(causal forest,CF)[4]等。其中,CF作为高维数据集下估计HTE的常用方法,已广泛应用于多个领域[5-8]。然而,这些方法仅适用于二元处理问题,而在现实世界中,多元处理情境更为普遍。例如,在实际临床中,医生通常根据治疗指南为目标患者推荐两到三种甚至更多的替代疗法[9];在营销活动中,商家可能提供多个策略,如广告投放、打折促销、跨店满减、会员专享及其组合等;在政策制定中,政府针对某一问题往往需要在多种政策之间做出选择,以实现最佳的社会效益。

为估计多元处理情境下的异质性因果效应,已有部分研究基于CF提出了多臂因果森林(multiarmed causal forest,MA-CF)[10]和多元处理-工具变量树(multiple treatment-instrumental variable tree, MT-IVT)[11]方法。然而,MA-CF仅限于在RCT数据中应用,MT-IVT则难以选择理想的工具变量。除了对CF方法的扩展,研究人员还开发了其他的多元处理效应估计方法,如基于回归的方法[12-14]、基于双重稳健的方法[15-17]以及基于BART的方法[18-19]。然而,基于回归和双重稳健的方法往往依赖于处理模型和结果模型的正确指定,当模型指定错误时,其性能往往表现不佳[20]。此外,基于BART的方法在处理高维数据时仍存在一定的局限性。

基于此,本文提出一种两阶段多元处理-因果森林(multiple treatment-causal forest,MT-CF)方法,旨在系统地估计并比较多种处理的异质性因果效应。第1阶段,利用传统CF方法分析成对的处理效应差异;第2阶段,将处理效应简化为一种特定形式,使之能够直接比较任一处理与所有处理的平均效应的差异,从而为多元处理场景下的异质性因果效应评估提供兼具高精度和强泛化性的估计方法。

1 多元处理潜在结果框架

本文依据Neyman-Rubin的潜在结果框架[21]估计多元处理的异质性因果效应。假设有 N 个个体构建的数据集 $D=\{\boldsymbol{Z}_i=(\boldsymbol{X}_i,W_i,Y_i)\}_{i=1}^{N}$,其中 $\boldsymbol{X}_i$、W_i、Y_i 分别表示第 i 个个体的特征向量、处理变量及结果变量。假设处理变量的取值有 m 种,每个个体只能接受其中一种处理,即如果个体 i 接受了处理 w,则 $W_i=w, w\in\{1,2,\cdots,m\}$。$T_{iw}(W_i)$ 表示个体 i 是否接受处理 w,定义为

$$T_{iw}(W_i)=\begin{cases}1, W_i=w\\0, \text{其他}\end{cases} \tag{1}$$

基于潜在结果框架,用 Y_{iw} 表示第 i 个个体接受处理 w 的潜在结果,那么对于任意两种不同的处理 j 和 k,$Y_{ij}-Y_{ik}$ 表示个体 i 接受处理 j 相对于接受处理 k 的个体处理效应(individual treatment effect, ITE)。由于只能观察到个体 i 所有潜在结果(Y_{i1},$\cdots$,Y_{im})中的一个,即 $Y_i=\sum_{w=1}^{m}T_{iw}(W_i)Y_{iw}$,个体 i 的ITE无法估计。因此,实践中的目标通常是估计条件平均处理效应(conditional average treatment effects, CATE)。处理 j 相对于处理 k 的CATE定义为

$$\tau_{j,k}(x)=E[Y_{ij}-Y_{ik}\mid\boldsymbol{X}_i=x] \tag{2}$$

式中 $E[\cdot]$ 为数学期望。

为了依据观测数据 $D=\{\boldsymbol{Z}_i=(\boldsymbol{X}_i,W_i,Y_i)\}_{i=1}^{N}$ 估计 $\tau_{j,k}(x)$,本文依赖于多元处理潜在结果框架的3个基本假设。

1)稳定的单位处理值假设,即确保一个个体的结果不受其他个体处理分配的影响(无干扰)。

2)弱混杂性假设,意味着个体 i 是否接受处理 w 不受潜在结果 Y_{iw} 的影响,而只受到特征向量 $\boldsymbol{X}_i$ 的影响,表示为

$$Y_{iw}\perp T_{iw}(W_i)\mid\boldsymbol{X}_i,\ \forall w\in\{1,2,\cdots,m\} \tag{3}$$

3)正值假设,意味着给定 $\boldsymbol{X}_i=x$ 的条件下,个体进入任一处理组的概率 P(即倾向得分)严格介于0~1之间,表示为

$$0<P(W_i=w\mid\boldsymbol{X}_i=x)<1,\ \forall i\in\{1,\cdots,N\},\ \forall w\in\{1,\cdots,m\} \tag{4}$$

2 MT-CF

相较于二元处理问题中判断是否施加处理的简单需求,多元处理情境下的核心在于确定对特定个体或亚组可能带来最佳结果的处理。尽管现有的处理效应估计方法在分析二元处理的异质性因果效应方面表现优异,但在多种处理情况下,选择接受哪种处理的样本作为实验组或对照组尚不清楚。通过将处理效应估计方法应用于每对处理可以解决此问题,然而,这种方法带来了巨大的解释困难。例如,若存在10种处理,则需要进行 $C_{10}^{2}=45$ 对成对比较。

虽然这种方法识别了每对处理的相对处理效应，但依旧无法明确特定个体的最佳处理。此外，在利用观测性数据估计处理效应时，若影响个体处理分配的因素与结局相关，则可能引入选择偏差。因此，亟需一种能够有效避免这些问题的新方法。

本文提出一种基于 CF 的多元处理效应估计方法 MT-CF，该方法分为成对处理效应估计和多元处理效应估计两个阶段。第 1 阶段通过提取对应成对处理的样本子集训练 CF 模型，估计成对处理之间的效应差异；第 2 阶段利用第 1 阶段训练的 CF 模型构建 MT-CF，将处理效应差异简化为一种形式，使其能估计每种处理与所有处理的平均效应差异，进而确定特定个体的最佳处理。图 1 展示了 MT-CF 方法的整体架构。

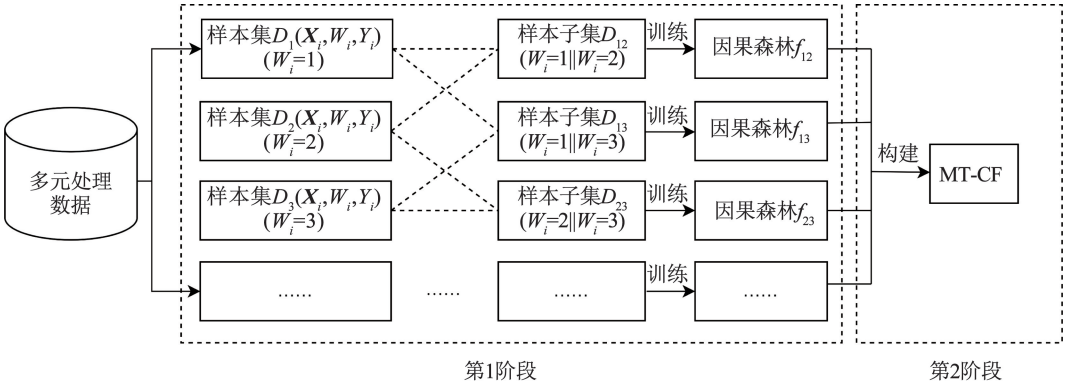


图 1　MT-CF 方法架构

Fig. 1　Framework of MT-CF

2.1　成对处理效应估计

第 1 阶段采用非参数诚实 CF 方法分析成对处理之间的效应差异，原因有以下 3 点：1）该方法递归地将样本划分为不同的群体，使得同一群体中的个体在不同处理之间表现出相似的结果差异，而不同群体中的个体则表现出不同的结果差异，从而识别最佳受益群体；2）基于树的方法因其非参数性质和数据驱动的建模方式，能够有效避免回归模型指定错误带来的偏差；3）CF 通过倾向得分匹配方法消除选择偏差的影响[22]，可以有效地利用观测性数据进行训练。

为了估计成对处理之间的效应差异，第 1 阶段为任意两种不同的处理构建对应的 CF。具体来说，对于第 j 种处理和第 k 种处理，首先从数据集 D 中提取样本集 $D_j=\{D\mid W_i=j\}$ 和 $D_k=\{D\mid W_i=k\}$。然后，将两者合并形成样本子集 $D_{jk}=\{D_j\cup D_k\}$。最后，利用该合并数据集构建 CF 的模型 f_{jk}，用于估计第 j 种处理相对于第 k 种处理的效应差异。

构建的 CF 模型由多棵因果树构成，每棵树使用叶子节点中处理组和对照组样本之间的平均结果差异估计 CATE。也就是说，给定样本 S，树 Π_b 中第 j 种处理相对于第 k 种处理的效应差异为

$$\hat{\tau}_b(x;S,\Pi_b)=\frac{1}{N_j^L}\sum_{i:\boldsymbol{X}_i\in L(x;\Pi_b)}Y_{ij}-\frac{1}{N_k^L}\sum_{i:\boldsymbol{X}_i\in L(x;\Pi_b)}Y_{ik} \tag{5}$$

式中：$L(x;\Pi_b)$ 为 $x\in\Pi_b$（第 b 棵树）中的某个叶子节点，N_j^L 和 N_k^L 分别为接受处理 j 和 k 的所有个体数量。

为了防止模型将特征和结果之间的虚假相关性（如过拟合）识别为处理的因果效应，CF 采用“honest”的方法，通过将数据分为两个样本集来最大化效应：一个用于分裂（S^{tr}），另一个用于叶子节点的效应估计（S^{est}）。因此，每棵因果树的分裂准则为

$$-\hat{R}_{\mathrm{EMSE}}(S^{\mathrm{tr}},N^{\mathrm{est}},\Pi_b)=\frac{1}{N^{\mathrm{tr}}}\sum_{i\in S^{\mathrm{tr}}}\hat{\tau}_b^2(\boldsymbol{X}_i;S^{\mathrm{tr}},\Pi_b)-\left(\frac{1}{N^{\mathrm{tr}}}+\frac{1}{N^{\mathrm{est}}}\right)\cdot\sum_{L\in\Pi_b}\left(\frac{S^2_{S_j^{\mathrm{tr}}}(L)}{p}+\frac{S^2_{S_k^{\mathrm{tr}}}(L)}{1-p}\right) \tag{6}$$

式中：N^{tr} 和 N^{est} 分别为训练样本和估计样本的大小，$S^2_{S_j^{\mathrm{tr}}}(L)$ 和 $S^2_{S_k^{\mathrm{tr}}}(L)$ 分别为叶子节点 L 中接受处理 j 和 k 的训练样本的方差，p 为训练样本接受处理 j 的概率。

单棵树的结果可能受样本划分限制，因此通过训练 B 棵树并综合所有树的估计，可以将个体 i 选择第 j 种处理相对于其选择第 k 种处理的效应差异表示为

$$\hat{\tau}_{j,k}(x)=\frac{1}{B}\sum_{b=1}^{B}\hat{\tau}_b(x),\ \forall j\neq k \tag{7}$$

第 1 阶段的具体细节见算法 1。

算法 1 成对处理效应估计

输入:$D=\{\boldsymbol{Z}_i=(\boldsymbol{X}_i,W_i,Y_i)\}_{i=1}^{N}$;处理变量个数 m;超参数设置:用于训练每棵树的样本比例 r,样本特征空间 $\boldsymbol{V}$,训练每棵树的停止准则

for $j=1,\cdots,m$ **do**
  for $k=1,\cdots,m$ **where** $j\neq k$ **do**
    从 D 中采样得到D_{jk}
    function $f_{jk}(D_{jk})$
      for $b=1,\cdots,B$ **do**
        以比例 r 从D_{jk}中采样得到子样本S_b和子样本空间$\boldsymbol{V}_b$以训练树 b
        将S_b划分为训练集S_b^{tr}和估计集S_b^{est}
        function $\Pi(S_b^{\mathrm{tr}},S_b^{\mathrm{est}},\boldsymbol{V}_b)$
          如果满足停止准则,则返回函数。否则执行以下语句:
          创建初始节点θ_0
          根据式(6)确定节点的拆分规则
          利用拆分规则将θ_0拆分为左节点θ_{left}和右节点θ_{right}
          将θ_{left}作为初始节点继续拆分生成新的节点
          将θ_{right}作为初始节点继续拆分生成新的节点
        end function
      保存树模型Π_b
    end for
      保存森林模型f_{jk}
    end function
  end for
end for

输出:$f_{jk};j=1,\cdots,m;k=1,\cdots,m;j\neq k$

2.2 多元处理效应估计

虽然第 1 阶段识别了每对处理的相对处理效应,但其结果在处理数量较多时难以解释,无法从中明确特定个体的最佳处理。为了解决这个问题,在第 2 阶段将处理效应构造为一种简洁的形式,使之能够直接比较任一处理与所有处理平均效应间的差异。具体来说,依据第 1 阶段得到的成对处理效应的差异,得到处理 j 相对于所有处理平均效应的差异,表达式为

$$\tau_{j,\mathrm{TA}}(x)=\frac{1}{B}\sum_{b=1}^{B}E\left[Y_{ij}-\frac{Y_{i1}+\cdots+Y_{im}}{m}\right]=\frac{1}{m}\frac{1}{B}\sum_{b=1}^{B}\sum_{k\neq j}E[Y_{ij}-Y_{ik}]=\frac{1}{m}\sum_{k\neq j}\tau_{j,k}(x) \tag{8}$$

第 2 阶段的具体细节见算法 2。

算法 2 多元处理效应估计

输入:$f_{jk};j=1,\cdots,m;k=1,\cdots,m;j\neq k$。测试样本特征向量$\boldsymbol{X}_i^{\mathrm{test}}$

for $j=1,\cdots,m$ **do**
  for $k=1,\cdots,m$ **where** $j\neq k$ **do**
    利用模型f_{jk}估计处理效应$\tau_{j,k}(\boldsymbol{X}_i^{\mathrm{test}})$
  end for
    根据式(8)估计$\tau_{j,\mathrm{TA}}(\boldsymbol{X}_i^{\mathrm{test}})$
    保存结果$\tau_{j,\mathrm{TA}}(\boldsymbol{X}_i^{\mathrm{test}})$
end for

输出:多元个体处理效应数组$[\tau_{1,\mathrm{TA}}(\boldsymbol{X}_i^{\mathrm{test}}),\tau_{2,\mathrm{TA}}(\boldsymbol{X}_i^{\mathrm{test}}),\cdots,\tau_{m,\mathrm{TA}}(\boldsymbol{X}_i^{\mathrm{test}})]$

3 实验研究

3.1 模拟实验

3.1.1 评价指标

为评估模型的性能,本文选取适用于连续型结果变量的评价指标,即标准误R_{SE}、平均绝对偏差R_{MAB}和均方根误差R_{RMSE},计算式分别为:

$$R_{\mathrm{SE}}=\sqrt{\frac{\sum_{i=1}^{N}(\hat{\tau}(x_i)-\bar{\tau}(x_i))^2}{N(N-1)}} \tag{9}$$

$$R_{\mathrm{MAB}}=\frac{1}{N}\sum_{i=1}^{N}|\hat{\tau}(x_i)-\tau(x_i)| \tag{10}$$

$$R_{\mathrm{RMSE}}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{\tau}(x_i)-\tau(x_i))^2} \tag{11}$$

由于采用蒙特卡罗模拟对模型性能进行检验,因此,最终评价结果取所有模拟运行指标的平均值。

3.1.2 蒙特卡罗模拟实验

MT-CF 方法的核心在于将成对处理的效应差异扩展至多元处理,因此在第 1 阶段中,二元处理估计方法(基学习器)是可替换的。本文设计了两种情景进行模拟,分别评估 MT-CF 方法相对于其他多元处理效应估计方法的性能以及在更换基学习器的情况下,MT-CF 方法相较于其他方法的性能表现。每种情景下,蒙特卡罗模拟进行的次数 $H=500$,每次模拟生成处理个数 $m=3$,样本个数 $N=1\ 000$。在每次模拟中,分别估计各处理相对于所有处理平均效应的差异。

从多元高斯分布$(X_{1i},\cdots,X_{10i})\sim N(0,\Sigma)$中生成 $d=10$ 个特征,真实倾向得分 P 是由多项式 logit(以处理水平 1 为基础水平)得到的[23],其计算式为:

$$P[W_i=1\mid\boldsymbol{X}_i=x]=\frac{1}{q_i} \tag{12}$$

$$P[W_i=2\mid\boldsymbol{X}_i=x]=\frac{\mathrm{e}x_{1i}}{q_i} \tag{13}$$

$$P[W_i=3|\boldsymbol{X}_i=x]=\frac{\mathrm{e}x_{2i}}{q_i} \tag{14}$$

$$\mathrm{e}x_{1i}=\exp\{1.5(-0.2+X_{1i}+X_{2i})\} \tag{15}$$

$$\mathrm{e}x_{2i}=\exp\{1.2(-0.1+X_{1i}+X_{2i})\} \tag{16}$$

$$q_i=1+\mathrm{e}x_{1i}+\mathrm{e}x_{2i} \tag{17}$$

式中q_i为归一化因子，其作用是在给定$\boldsymbol{X}_i=x$的条件下，确保个体进入不同处理组的概率之和为 1。

给定概率 P 和[0,1]内的均匀随机变量μ_i，W_i的计算式为

$$W_i=\begin{cases}1,\mu_i\leqslant P[W_i=1|\boldsymbol{X}_i=x]\\2,P[W_i=1|\boldsymbol{X}_i]<\mu_i\leqslant P[W_i=1|\boldsymbol{X}_i]+P[W_i=2|\boldsymbol{X}_i=x]\\3,\text{其他}\end{cases} \tag{18}$$

$\tau(x)$和结果Y_i的计算式分别为：

$$\tau(x)=\frac{1}{1+\exp(-X_{3i})} \tag{19}$$

$$Y_i=\max(X_{2i}+X_{3i},0)+\frac{X_{4i}+X_{5i}+X_{6i}}{2}+\tau(x)W_i+\varepsilon_i,\varepsilon_i\sim N(0,1) \tag{20}$$

式中ε_i为误差项。

在第 1 种情景中，将基于 CF 的 MA-CF[10]、基于回归的传统回归调整（regression adjustment，RA）[12]、基于双重稳健的目标最大似然估计（targeted maximum likelihood estimation，TMLE）[17]以及基于 BART 的随机截距加速失效时间模型（random-intercept accelerated failure time-Bayesian additive regression trees，riAFT-BART）[19]作为基线与 MT-CF 进行比较。

在第 2 种情景中，P、W_i、$\tau(x)$和Y_i是通过提取C_3^2个不同二元类别的子样本生成的。基学习器包括两种：一种是二元因果机器学习方法，即 BART[2]；另一种是传统的机器学习方法，将处理作为输入特征获得预测结果，包括线性回归（linear regression，LR）[24]、支持向量机（support vector machine，SVM）[25]和极端梯度提升（extreme gradient boosting，XGBoost）[26]方法。以上方法均利用式（8）进行 CATE 估计。

表 1 和表 2 分别为两种情景下蒙特卡罗模拟的详细结果。由表 1 可知，与其他多元处理效应估计方法相比，MT-CF 在平均绝对偏差和均方根误差方面的表现均具有显著的优越性。这是由于 MT-CF 方法通过成对估计处理效应差异并整合结果，充分利用了 CF 的精细化建模能力，降低了干扰和偏差，从而提高了估计精度。

表 1　多元处理效应估计方法比较

Tab. 1　Comparison of multiple treatment effect estimation methods

方法	$W_i=1$			$W_i=2$			$W_i=3$		
	R_{SE}	R_{MAB}	R_{RMSE}	R_{SE}	R_{MAB}	R_{RMSE}	R_{SE}	R_{MAB}	R_{RMSE}
MT-CF	0.003	**0.336**	**0.393**	**0.002**	**0.125**	**0.154**	**0.004**	**0.376**	**0.387**
MA-CF	0.003	0.359	0.417	0.002	0.131	0.158	0.004	0.390	0.405
RA	0.006	0.457	0.541	0.009	0.364	0.449	0.011	0.527	0.554
TMLE	**0.002**	0.346	0.406	0.002	0.129	0.162	0.004	0.384	0.393
riAFT-BART	0.004	0.392	0.427	0.003	0.232	0.274	0.005	0.405	0.406

表 2　多元处理–基学习器比较

Tab. 2　Comparison of multiple treatment methods with base learners

方法	$W_i=1$			$W_i=2$			$W_i=3$		
	R_{SE}	R_{MAB}	R_{RMSE}	R_{SE}	R_{MAB}	R_{RMSE}	R_{SE}	R_{MAB}	R_{RMSE}
MT-CF	**0.003**	**0.336**	**0.393**	**0.002**	**0.125**	**0.154**	**0.004**	**0.376**	**0.387**
MT-BART	0.004	0.357	0.434	0.004	0.177	0.197	0.009	0.424	0.433
MT-LR	0.005	0.507	0.549	0.004	0.180	0.216	0.006	0.428	0.483
MT-SVM	0.004	0.506	0.548	0.003	0.184	0.221	0.007	0.495	0.537
MT-XGBoost	0.007	0.380	0.474	0.006	0.231	0.282	0.012	0.385	0.448

由表 2 可知，在更换基学习器后，MT-CF 仍展现出优于其他方法的性能。尽管 CF 和 BART 都是非参数方法，但 BART 对先验分布依赖性较强且对先验选择敏感，不当选择可能影响模型表现。相比之下，CF 无需对数据分布形态或函数形式做出强假设，其通过构建大量因果树并进行“honest”估计，在

因果关系建模中展现出更高的准确性。而传统机器学习方法由于仅估计相关性而非因果性,难以有效估计处理效应间的真实差异,因而其性能相对较差。

3.2 实例研究

近年来,乳腺癌发病率在全球范围内持续上升,早期治疗已被广泛认可为降低乳腺癌死亡率的重要措施[27]。本文以监测、流行病学和最终结果(the surveillance, epidemiology, and end results, SEER)数据库[28]中 14 300 位接受改良根治性乳房切除术(the modified radical mastectomy, MRM)的早期浸润性(Ⅰ、Ⅱa、Ⅱb)乳腺癌患者为例进行实例研究,探究术后辅助治疗对患者生存期的异质性因果影响,从而检验 MT-CF 方法在实际应用中的合理性及有效性。具体过程如图 2 所示。

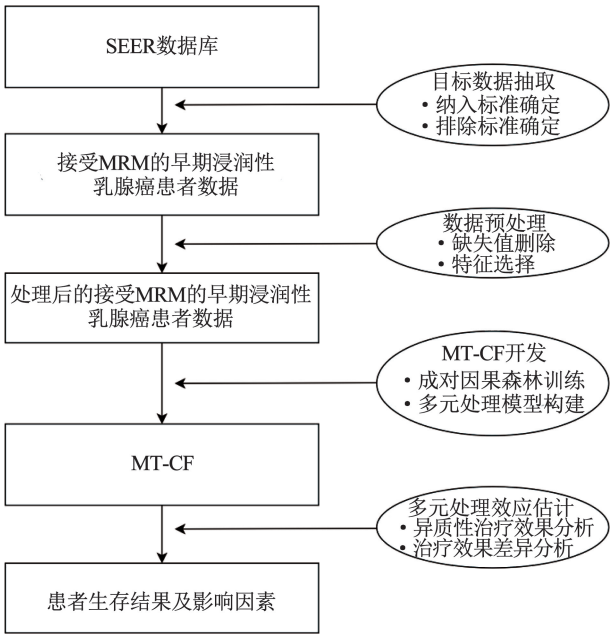


图 2 乳腺癌术后辅助治疗疗效估计过程

Fig. 2 Estimation of postoperative adjuvant therapy efficacy in breast cancer

3.2.1 数据获取与处理

本文从 SEER 数据库中获取诊断年份为2006—2015 年之间接受 MRM 的早期浸润性乳腺癌病例数据,共筛选出 14 300 例样本纳入研究。

在对样本进行缺失值处理和特征选择之后,最终用于构建 MT-CF 的特征,包括家庭收入水平、诊断年龄、种族、病理类型、组织学分级、TNM 分期、阳性区域淋巴结数量、肿瘤大小以及淋巴结受累信息。其中,收入根据中位数划分为低收入(≤54 999 美元)、中等收入(54 999 ~ 70 000 美元)和高收入(≥70 000 美元)。

依据对复发风险的个体化评估,术后辅助治疗的选择包括化疗、放疗、系统治疗或联合治疗[29-33],即:“辅助化疗 + 辅助系统治疗”(治疗方案 1)、“辅助化疗 + 辅助放疗 + 辅助系统治疗”(治疗方案 2)、“仅辅助放疗”(治疗方案 3)、“无辅助治疗”(治疗方案 4)、“辅助化疗 + 辅助放疗”(治疗方案 5)、“辅助放疗 + 辅助系统治疗”(治疗方案 6)、“仅辅助系统治疗”(治疗方案 7)和“仅辅助化疗”(治疗方案 8)。主要结局为总生存期(overall survival, OS),定义为患者从首次诊断为早期浸润性乳腺癌至因任何原因死亡所经历的时间间隔(月)。

3.2.2 实验设计

在实际临床中,医生通常依赖经验和知识制定治疗方案,往往关注术后辅助治疗在人群中的平均治疗效果,这使得个体化治疗较为困难。为解决这一问题,本文提出的 MT-CF 方法考虑了辅助治疗效果的异质性,能够评估患者的异质性治疗效果,进而为患者推荐更精准的治疗方案。本文设计了两个实验以检验 MT-CF 方法在实际应用中的有效性。1)分析不同特征患者在不同治疗方案间的效果差异是否显著,评估 MT-CF 识别异质性处理效应的能力。2)探究异质性治疗效果与平均治疗效果的差异,验证所提方法的有效性。

为了确定在某些患者组别上与所有治疗的平均效果有显著统计学差异的治疗方案,首先为每对治疗方案构建 CF 模型,共需要 $C_8^2 = 28$ 个森林。对于每位患者,估算出任一治疗与所有治疗的平均效果之间的 OS 差异,并利用 Bootstrap 方法估算出差异的标准误差。为了直观地说明不同治疗方案对不同患者治疗效果的异质性,根据影响患者治疗效果差异的重要因素(肿瘤大小、年龄和家庭收入水平)对患者进行了分组。对于每个患者组,采用 $V_{igk} \in \{-1,0,1\}$ 表示治疗方案 k 是否在 10% 的显著性水平上显著低于、等于或优于组 g 中患者 i 的所有治疗方案的平均效果。然后,通过计算 $\bar{V}_{gk} = 1/N_{gk} \sum_{i=1}^{N_{gk}} V_{igk}$ 评估患者组 g 的治疗方案 k 的整体表现,其中 N_{gk} 表示治疗方案为 k 的组 g 中的患者数量。

为解决因变量 OS 在零值处存在左截断导致的传统回归偏差问题,采用 Tobit 模型对平均治疗效果进行建模,并基于模型计算结果选择最佳治疗方案。在模型中,采用按治疗方案聚类的稳健标准误差,以考虑任意组内相关性导致的方差或标准误差的差异[34]。在所有患者组别中,OS 最大的治疗方案被指定为最好的治疗方案。为了获得考虑辅助治疗异质性的最佳方案,本文将 MT-CF 方法应用于不同家庭收入水平的所有患者。

3.2.3　实验结果及讨论

表 3 以 6 位患者为例，总结了患者选择不同治疗方案的效果差异（从左至右依次为第 1 ~ 6 位）。可以注意到，有些治疗方案（如治疗方案 1）对所有 6 位患者的治疗效果均优于平均效果，而有些治疗方案（如治疗方案 8）对大多数患者的治疗效果却不如平均效果。然而，对于比平均效果好（或差）的治疗方案来说，个别病人的差异程度是不同的。例如，第 2 位患者采用治疗方案 1，疗效较平均水平提升了 2.89 个月；第 4 位患者采用治疗方案 1，疗效较平均水平提升了 10.34 个月。此外，有些治疗方案对某些病人的疗效优于平均效果，但对另一些病人的疗效却差于或等于平均效果。例如，治疗方案 6 对第 1 位和第 4 位患者的治疗效果较好，但对第 3 位、第 5 位和第 6 位患者的治疗效果较差，而对第 2 位患者，其治疗效果和平均效果没有显著差异。这些数据表明，不同患者接受不同治疗方案的治疗效果差异是异质的。

表 3　不同患者接受单一治疗的效果与所有治疗的平均效果比较

Tab. 3　Comparison of the effects of a single treatment for different patients with the average effect of all treatments

治疗方案	OS 差异量/月					
	0 ~ <10 mm，43 岁，高收入	10 ~ <20 mm，68 岁，低收入	20 ~ <30 mm，42 岁，中等收入	30 ~ <40 mm，83 岁，高收入	40 ~ <50 mm，78 岁，中等收入	50 ~ <60 mm，49 岁，低收入
1	3.03***	2.89***	**7.98***	**10.34***	**6.80***	**6.67***
2	−0.94***	4.26***	0	4.77***	4.80***	4.59***
3	**11.76***	**9.87***	3.47***	−4.56***	−2.81***	3.94***
4	2.13***	4.02***	0.80***	−6.84***	−9.28***	−3.20***
5	−8.18***	−4.14***	−6.23***	−0.87***	1.45***	−0.38***
6	2.65***	−0.04	−1.48***	3.21***	−1.57***	−9.03***
7	−3.29***	−0.58***	−0.14*	−0.17**	0.62***	−3.35***
8	−6.78***	−16.06***	−4.40***	−6.21***	−0.02	−0.09

注：*、** 和 *** 分别表示治疗方案在 10%、5% 和 1% 的显著性水平上优于或劣于所有治疗的平均效果。

图 3 为不同治疗方案在按肿瘤大小、年龄（每个区间内 4 列依次为 40 ~ 80 岁的 4 个年龄组）和家庭收入水平分组的患者中的表现，可以看出，顶部行中的大多数单元格（如与治疗方案 1 ~ 3 相关的单元格）的颜色表明这些治疗方案在大多数患者中的表现优于所有治疗的平均效果。相比之下，底部各行（如与治疗方案 6 ~ 8 相关的单元格）的颜色表示这些治疗方案在大多数患者中的表现劣于所有治疗的平均效果。顶端行中有混合颜色的单元格，表示治疗方案 1 对大部分患者群体的治疗效果较好，但对于非高收入且肿瘤较小的中年患者群体，该方案的治疗效果与平均水平没有显著差异。同样，接近底部的行中有混合颜色的单元格，表示这些治疗方案对大部分患者群体的治疗效果较差，但对某些患者群体，其治疗效果与平均水平没有显著差异。此外，有些治疗方案（如治疗方案 2 ~ 7）在某些患者组别方面明显优于所有治疗的平均效果，而在其他患者组别方面则明显劣于所有治疗的平均效果。因此，对于具有不同特征的患者，治疗方案之间的效果差异是异质的。

从图 3 也可以得到一些额外的信息。例如，当选择治疗方案 4（即不治疗）时，当肿瘤大小超过 20 mm时，随着肿瘤的增大，其治疗效果显著低于平均效果。这一发现与现有研究中的结论相一致，即肿瘤大于 20 mm 与死亡率升高有关[35]。此外，当选择治疗方案 3（辅助放疗）时，患者的总生存期得以显著提高。选择治疗方案 7（辅助系统治疗）或治疗方案 8（辅助化疗）的患者生存状况更差，这可能与药物的多种不良反应有关[36-38]。联合治疗的效果则提供了更深入的见解。当患者选择治疗方案 2（辅助化疗 + 辅助放疗 + 辅助系统治疗）时，尽管某些患者可能因药物副作用而导致疗效劣于平均效果，但总体上显示出优于单一治疗的效果。这表明联合治疗可能通过多重机制发挥作用，从而在控制肿瘤进展方面提供了更大的优势。

表 4 总结了最佳治疗方案的集合（第 2 列）以及根据平均治疗效果（第 3 列）和异质性治疗效果（第 4 列）得出的对应每种最佳治疗方案的患者人数。

从表 4 可以看出，当从平均治疗效果转向异质性治疗效果时，最佳治疗方案及其对应的患者 OS（第 5 列）会出现差异，说明在寻找最佳治疗方案时考虑治疗异质性的必要性。MT-CF 方法不仅能引导患者选择能够提高其预期 OS 的治疗方案，还能将患者引导至更多元化的治疗方案选择上。从医疗

能力的角度来看,这能更容易地为患者提供最佳治疗。表 4 提供了两个主要信息。1)在整个人群中,平均治疗效果最好的治疗方案对将近一半(约 45%)的患者来说效果并不是最好的。基于治疗异质性的信息显示,就低收入水平患者而言,采用平均治疗效果确定的最佳治疗方案仅是少数患者的首选方案。2)根据患者异质性治疗效果而非平均治疗效果来选择最佳治疗方案会大幅提高总体平均 OS。对于 3 种收入水平的患者而言,低收入患者的 OS 提高幅度最大,而高收入患者的 OS 提高幅度最小。

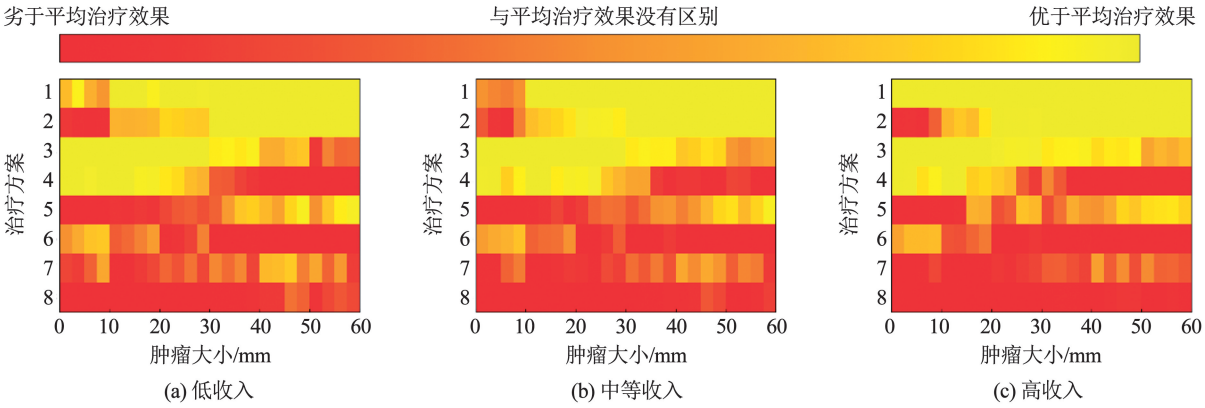


图 3 不同患者群接受单一治疗的效果与所有治疗的平均效果比较

Fig. 3 Comparison of the effects of a single treatment received by different patient groups with the average effect of all treatments

表 4 基于异质性治疗效果的患者 OS 提高情况

Tab. 4 Improvement in overall survival of patients based on heterogeneous treatment effects

收入水平	治疗方案	患者数量		OS 提高量/月	
		平均治疗效果	异质性治疗效果	均值	标准误差
低收入	1	852	270	0	0
	2		20	1.59***	0.36
	3		524	5.68***	0.14
	4		23	3.17***	0.29
	5		4	0.71*	0.29
	6		7	0.73***	0.12
	7		4	1.00**	0.19
	总体			3.63***	0.13
中等收入	1	1 112	540	0	0
	2		14	0.91**	0.31
	3		496	5.57***	0.18
	4		55	3.85***	0.27
	5		4	1.46***	0.14
	6		3	1.24	0.82
	总体			2.69***	0.11
高收入	1	896	763	0	0
	2		4	0.39	0.19
	3		110	4.44***	0.32
	4		19	1.21***	0.27
	总体			0.57***	0.06

注:*、** 和 *** 分别表示 OS 提高量在 10%、5% 和 1% 的水平上显著。

图 4 为从平均治疗效果确定的最佳方案转换为异质性治疗效果确定的最佳方案时,按肿瘤大小、年龄和家庭收入水平分组的患者 OS 的平均提高情况。其中,各区间内 4 个点对应图 3 的 4 个年龄组,

实心圆点和空心圆点分别表示在 10% 的水平上，OS 平均提高量是显著和不显著。由图 4 可以看出，对于低收入患者，无论肿瘤大小，获益均较多。而对于中高收入患者而言，肿瘤较小（40 mm 以内）时获益较多。对本研究中的患者而言，其意义在于，考虑异质性治疗效果而不是平均治疗效果，可以使很大一部分患者更准确地确定最适合自己的治疗方案。

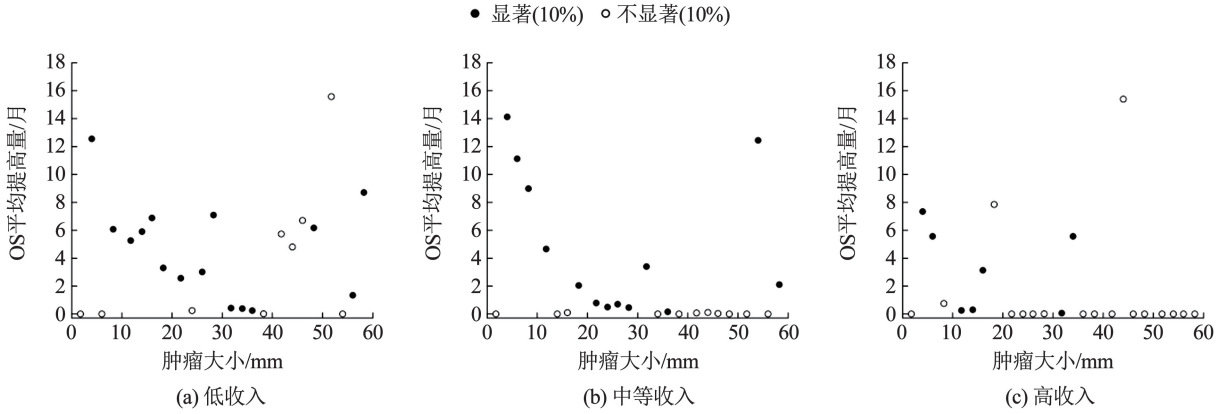


图 4　不同患者组别的 OS 平均提高量

Fig. 4　Average improvement in OS of different patient groups

根据患者的异质性治疗效果选择治疗方案，可以显著提升治疗效果并延长患者的生存期。对于临床医生而言，这种方法有助于为患者量身定制更精准的治疗策略，从而增强医疗决策的科学性和可靠性。医疗机构也能够通过个性化方法更有效地配置资源，减少不必要的医疗开支，同时提高整体医疗服务的质量和效率。

4　结　论

本文将传统 CF 方法拓展至多元处理情境，提出两阶段 MT-CF 方法。通过两组模拟实验以及实例研究，验证了所提方法的有效性。主要结论如下：

1）利用传统 CF 估计成对处理的效应差异，然后将成对处理的因果效应差异扩展至多元处理，从而间接地估计每种处理的处理效应。基于树的方法有效减轻了回归模型指定错误带来的偏差，并能有效识别在不同处理间表现出显著效应差异的组别。

2）MT-CF 方法有效提升了多元处理效应估计的精度。对比实验表明，在两种情景下，MT-CF 的整体性能均明显优于其他基准方法。

3）在 SEER 乳腺癌数据集的实验中，MT-CF 有效地比较了多种治疗方案的疗效差异，明确了不同亚组的最佳治疗方案，展示了该方法在实际医疗数据中的广泛应用潜力。

参考文献

[1] Zabor E C, Kaizer A M, Hobbs B P. Randomized controlled trials[J]. Chest, 2020, 158(1): S79. DOI: 10.1016/j.chest.2020.03.013

[2] Chipman H A, George E I, McCulloch R E. BART: Bayesian additive regression trees[J]. Annals of Applied Statistics, 2010, 4(1): 266. DOI: 10.1214/09-AOAS285

[3] Athey S, Imbens G. Recursive partitioning for heterogeneous causal effects[J]. Proceedings of the National Academy of Sciences of the United States of America, 2016, 113(27): 7353. DOI: 10.1073/pnas.1510489113

[4] Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests[J]. Journal of the American Statistical Association, 2018, 113(523): 1228. DOI: 10.1080/01621459.2017.1319839

[5] Ngufor C, Yao Xiaoxi, Inselman J W, et al. Identifying treatment heterogeneity in atrial fibrillation using a novel causal machine learning method[J]. American Heart Journal, 2023, 260: 124. DOI: 10.1016/j.ahj.2023.02.015

[6] Kreif N, DiazOrdaz K, Moreno-Serra R, et al. Estimating heterogeneous policy impacts using causal machine learning: a case study of health insurance reform in Indonesia[J]. Health Services and Outcomes Research Methodology, 2022, 22(2): 192. DOI: 10.1007/s10742-021-00259-3

[7] 王喆，毛倩，管佩霞，等. 基于因果森林的三阴性乳腺癌术后放疗的个性化疗效评价[J]. 中国卫生统计，2023，40(1)：32
Wang Zhe, Mao Qian, Guan Peixia, et al. Assessment of individualized treatment effect of postoperative radiotherapy in patients with triple negative breast cancer based on causal forests[J]. Chinese Journal of Health Statistics, 2023, 40(1): 32. DOI: 10.11783/j.issn.1002-3674.2023.01.007

[8] Kakimoto S, Mieno T, Tanaka T S T, et al. Causal forest approach for site-specific input management via on-farm precision experimentation[J]. Computers and Electronics in Agriculture, 2022, 199: 107164. DOI: 10.1016/j.compag.2022.107164

[9] Flaig T W, Spiess P E, Agarwal N, et al. Bladder cancer, version 3.2020, NCCN clinical practice guidelines in oncology[J]. Journal of the National Comprehensive Cancer Network, 2020, 18(3): 329. DOI: 10.6004/jnccn.2020.0011

[10] Sverdrup E, Wu Han, Athey S, et al. Qini curves for multi-armed treatment rules[J/OL]. Journal of Computational and Graphical Statistics, 2024-12-10. https://doi.org/10.1080/10618600.2024.

2418820

[11] Wang Guihua, Li Jun, Hopp W J. Personalized healthcare outcome analysis of cardiovascular surgical procedures[J]. Manufacturing and Service Operations Management, 2023, 25(4): 1567. DOI: 10.1287/msom.2023.1227

[12] Rubin D B. Using multivariate matched sampling and regression adjustment to control bias in observational studies[J]. Journal of the American Statistical Association, 1979, 74(366): 318. DOI: 10.1080/01621459.1979.10482513

[13] Hong Guanglei. Marginal mean weighting through stratification: a generalized method for evaluating multivalued and multiple treatments with nonexperimental data[J]. Psychological Methods, 2012, 17(1): 44. DOI:10.1037/a0024918

[14] Robins J M, Rotnitzky A. Semiparametric efficiency in multivariate regression models with missing data[J]. Journal of the American Statistical Association, 1995, 90(429): 122. DOI: 10.2307/2291135

[15] Cattaneo M D. Efficient semiparametric estimation of multi-valued treatment effects under ignorability[J]. Journal of Econometrics, 2010, 155(2): 138. DOI:10.1016/j.jeconom.2009.09.023

[16] Uysal S D. Doubly robust estimation of causal effects with multivalued treatments: an application to the returns to schooling[J]. Journal of Applied Econometrics, 2015, 30(5): 763. DOI: 10.1002/jae.2386

[17] Rose S, Normand S L. Double robust estimation for multiple unordered treatments and clustered observations: evaluating drug-eluting coronary artery stents[J]. Biometrics, 2019, 75(1): 289. DOI:10.1111/biom.12927

[18] Hu Liangyuan, Gu Chenyang, Lopez M, et al. Estimation of causal effects of multiple treatments in observational studies with a binary outcome[J]. Statistical Methods in Medical Research, 2020, 29(11): 3218. DOI:10.1177/0962280220921909

[19] Hu Liangyuan, Ji Jiayi, Ennis R D, et al. A flexible approach for causal inference with multiple treatments and clustered survival outcomes[J]. Statistics in Medicine, 2022, 41(25): 4982. DOI: 10.1002/sim.9548

[20] Linden A, Uysal S D, Ryan A, et al. Estimating causal effects for multivalued treatments: a comparison of approaches[J]. Statistics in Medicine, 2016, 35(4): 534. DOI:10.1002/sim.6768

[21] Imbens G W, Rubin D B. Causal inference for statistics, social, and biomedical sciences: an introduction[M]. Cambridge: Cambridge University Press, 2015

[22] Athey S, Tibshirani J, Wager S. Generalized random forests[J]. Annals of Statistics, 2019, 47(2): 1148. DOI: 10.1214/18-AOS1709

[23] Cattaneo M D, Drukker D M, Holland A D. Estimation of multivalued treatment effects under conditional independence[J]. Stata Journal, 2013, 13(3): 407. DOI: 10.1177/1536867X1301300301

[24] Marill K A. Advanced statistics: linear regression, part II: multiple linear regression[J]. Academic Emergency Medicine, 2004, 11(1): 94. DOI:10.1197/S1069-6563(03)00601-8

[25] Cortes C, Vapnik V. Support-vector networks[J]. Machine Learning, 1995, 20(3):273. DOI:10.1023/A:1022627411411

[26] Chen Tianqi, Guestrin C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: Association for Computing Machinery, 2016: 785. DOI: 10.1145/2939672.2939785

[27] 管红娇, 张英涛, 唐降龙. 基于超声和钼靶的两阶段乳腺癌诊断系统[J]. 哈尔滨工业大学学报, 2019, 51(11): 8
Guan Hongjiao, Zhang Yingtao, Tang Xianglong. Two-stage breast cancer diagnosis system based on ultrasound and mammogram images[J]. Journal of Harbin Institute of Technology, 2019, 51(11): 8. DOI:10.11918/j.issn.0367-6234.201904005

[28] Park H S, Lloyd S, Decker R H, et al. Overview of the surveillance, epidemiology, and end results database: evolution, data variables, and quality assurance[J]. Current Problems in Cancer, 2012, 36(4): 183. DOI: 10.1016/j.currproblcancer.2012.03.007

[29] Rossi L, Stevens D, Pierga J Y, et al. Impact of adjuvant chemotherapy on breast cancer survival: a real-world population[J]. Plos One, 2015, 10(7): 1. DOI:10.1371/journal.pone.0132853

[30] Johnson M E, Handorf E A, Martin J M, et al. Postmastectomy radiation therapy for T3N0: a SEER analysis[J]. Cancer, 2014, 120(22): 3569. DOI:10.1002/cncr.28865

[31] Sieluk J, Yang Lingfeng, Haiderali A, et al. Systemic therapy, survival and end-of-life costs for metastatic triple-negative breast cancer: retrospective SEER-Medicare study of women age ⩾65 years[J]. Future Oncology, 2021, 17(20): 2581. DOI: 10.2217/fon-2021-0019

[32] Nielsen H M, Overgaard M, Grau C, et al. Study of failure pattern among high-risk breast cancer patients with or without postmastectomy radiotherapy in addition to adjuvant systemic therapy: long-term results from the Danish Breast Cancer Cooperative Group DBCG 82 b and c randomized studies[J]. Journal of Clinical Oncology, 2006, 24(15): 2268. DOI: 10.1200/JCO.2005.02.8738

[33] 刘昌娟. 探讨免疫治疗联合化疗对乳腺癌患者术后免疫功能和生存状况的影响[J]. 中外医疗, 2024, 43(14): 102
Liu Changjuan. Effect of immunotherapy combined with chemotherapy on postoperative immune function and survival of patients with breast cancer[J]. China & Foreign Medical Treatment, 2024, 43(14): 102. DOI:10.16662/j.cnki.1674-0742.2024.14.102

[34] Batt R J, Terwiesch C. Early task initiation and other load-adaptive mechanisms in the emergency department[J]. Management Science, 2017, 63(11): 3531. DOI:10.1287/mnsc.2016.2516

[35] Lim S H, Delaney G P, Descallar J, et al. Outcomes of ethnic minority groups with node-positive, non-metastatic breast cancer in two tertiary referral canters in Sydney, Australia[J]. Plos One, 2014, 9(4): e95852. DOI:10.1371/journal.pone.0095852

[36] Azim H A, De Azambuja E, Colozza M, et al. Long-term toxic effects of adjuvant chemotherapy in breast cancer[J]. Annals of Oncology, 2011, 22(9): 1939. DOI:10.1093/annonc/mdq683

[37] Wolff A C, Blackford A L, Visvanathan K, et al. Risk of marrow neoplasms after adjuvant breast cancer therapy: the national comprehensive cancer network experience[J]. Journal of Clinical Oncology, 2015, 33(4): 340. DOI:10.1200/JCO.2013.54.6119

[38] Hamelinck V C, Bastiaannet E, Pieterse A H, et al. Patients' preferences for surgical and adjuvant systemic treatment in early breast cancer:a systematic review[J]. Cancer Treatment Reviews, 2014, 40(8): 1005. DOI:10.1016/j.ctrv.2014.06.007