联邦生存森林

Alberto Archetti*DEIB* 米兰理工大学 米兰，意大利 al berto.archetti@polito.it

Matteo Matteucci*DEIB* 米兰理工大学米兰，意大利 matteo. matteucci@polimi.it

arXiv:2302.02807v2 [cs.LG] 7 Aug 2023

摘要—**生存分析是统计学的一个分支，旨在对某一特定事件在总体中发生的时间进行建模。生存分析已广泛应用于医疗保健、工程学和社会科学领域。然而，现实世界中的应用涉及分布式、不完整、删失且机密的生存数据集。在此背景下，联邦学习能够大幅提升生存分析应用的性能。联邦学习提供了一系列保护隐私的技术，可在不损害用户隐私的情况下，基于多个数据集联合训练机器学习模型，从而实现更好的泛化性能。然而，尽管近年来人工智能研究中联邦学习得到了广泛发展，关注联邦生存分析的研究仍然很少。在本文中，我们提出了一种用于生存分析的新型联邦算法，该算法基于最成功的生存模型之一——随机生存森林。我们将所提出的方法称为联邦生存森林（FedSurF）。仅通过一个通信轮次，FedSurF 就能获得与经过数百次联邦迭代训练的基于深度学习的联邦模型相当的判别能力。此外，FedSurF 保留了随机森林的所有优势，即计算成本低，并且能够自然地处理缺失值和不完整数据集。这些优势在现实的联邦环境中尤为理想，因为此类环境包含多个规模较小的数据集，且这些数据集存储在计算能力较低的设备上。数值实验将 FedSurF 与联邦网络中的先进生存模型进行了比较，结果表明，在具有非独立同分布数据的现实环境中，FedSurF 的性能优于基于深度学习的联邦算法。**

索引词—**深度学习、联邦学习、随机森林、生存分析**

I. 引言

生存分析 [1], [2],也称为事件发生时间分析，是统计学的一个分支，致力于对个体群体中某一关注事件发生前的时间进行建模。具体而言，生存模型基于统计学和机器学习技术，预测生存函数$S(t)$。该函数评估某个研究对象截至时间 t 尚未经历该事件的概率。生存分析已在医疗保健、工程学、经济学和社会科学应用中取得广泛成功 [2]。然而，在现实场景中，生存数据通常是分布式的、不准确的且不完整的 [3], [4]。此外，生存数据可能包含相当大比例的删失样本，即个体尚未经历该事件的数据记录。即使生存模型适合从删失样本中提取所有有用信息，数据集中的高删失比例仍然给生存模型带来了严峻挑战才能取得成功。一种可能的解决方案是增加可用于训练的数据样本数量。然而，大多数生存应用依赖机密数据，这些数据由于安全或监管约束而具有隐私性且无法共享 [3], [4]。

在此背景下，联邦学习（FL） [5], [6] 能够显著提升生存应用在现实场景中的成功率。FL是一种机器学习领域，其中一组客户端各自持有一个私有数据集，并在中央服务器的协调下协作训练机器学习模型。分布式学习与FL之间的关键区别在于，私有数据信息绝不会离开其采集和存储所在的设备。FL通过在客户端与中央服务器之间迭代交换模型参数，构建一个在联邦中对所有客户端都具有良好性能的平均模型。如此一来，FL从设计上允许多方共同训练机器学习模型，而不会泄露本地设备上存储的私有数据信息。联邦模型能够利用规模庞大且具有代表性的数据池进行训练，因此比其本地模型具有更好的泛化性能。

FL能够帮助生存分析克服生存数据稀缺、删失以及机密性带来的局限。为此，联邦生存分析领域研究将生存模型整合到联邦算法中的技术。在与联邦生存分析相关的研究中，大多数工作聚焦于Cox模型 [3], [7]–[16]。尽管Cox模型在经典生存分析中占据重要地位，但它建立在比例性假设之上，而这一假设在大规模联邦数据集中可能并不成立。其他研究则将联邦生存分析拓展至深度神经模型 [17]–[19]。尽管这些模型功能强大，其训练需要多个通信轮次，这可能会影响收敛速度和带宽使用量。据我们所知，集成学习在联邦生存分析中的应用尚未得到探索。

在本研究中，我们提出了一种新颖的联邦式改编方案，应用于机器学习文献中最成功的生存模型之一——随机生存森林（RSF） [20]。我们将所提出的算法称为联邦生存森林（FedSurF）。在 FedSurF 中，每个客户端在本地使用自己的数据训练生存森林，并向中央服务器发送构成本地森林的一组经过精心挑选的树。随后，中央服务器构建由客户端设备上训练的树组成的树集成。对于每个客户端，树选择可以随机进行

采用均匀概率，或依据本地评估进行。具体而言，我们讨论一种基于综合布里尔评分（IBS） [21]的树采样技术。

借助 FedSurF，我们旨在将 RSF 及其相较于最先进生存模型的优势引入联邦设置。具体而言，RSF 的计算需求低于深度学习模型，且所需的超参数调优较少，因此不易发生过拟合。其次，RSF 能自然地处理缺失值和分类变量，从而能够从不完整、不准确且经过删失的本地数据集中提取大部分有用信息。第三，由于采用基于树的结构，RSF 更易于解释。最后，就联邦训练而言，FedSurF 只需客户端与中央服务器之间进行一次通信轮次，即可构建最终模型。从节点间通信的角度看，这一特点使 FedSurF 比用于训练联邦深度学习模型的传统迭代算法高效得多。

FedSurF 已与若干最先进的生存模型一同在联邦场景中进行了测试。数值实验表明，所提出的算法能够有效解决生存问题，并具有很强的泛化能力。此外，在数据非独立同分布的联邦中开展的实验显示，FedSurF 能够适应现实场景，在这些场景中，每个客户端的数据分布各不相同。具体而言，在大多数此类场景中，FedSurF 的表现优于深度学习模型。

本文其余部分安排如下。第二节全面概述联邦学习和生存分析的背景及其最新进展。第三节介绍我们提出的算法 FedSurF。第四节详尽分析实验结果，以证明 FedSurF 相较于现有方法的有效性。文中详细说明了所有实验流程，并公开了源代码，以促进研究结果的可复现性。最后，第五节总结全文。

II. 背景与相关工作

本节概述生存分析和联邦学习的相关文献，重点介绍联邦生存分析领域当前的研究进展。

A. 生存分析

生存分析（又称事件发生时间分析）是统计学的一个分支，用于对某一群体中感兴趣事件的发生时间进行建模。生存问题的主要目标是估计事件发生概率随时间变化的情况。更正式地说，生存模型的输出是生存函数

$$S(t|\mathbf{x}) = P(T > t|\mathbf{x})$$

它用于评估某个个体截至时间 t 尚未经历该事件的概率。该个体由特征向量 $\mathbf{x} \in \mathbb{R}^d$ 表征。生存函数利用机器学习和统计技术，根据生存数据集估计。生存数据集D是一组三元组 $(\mathbf{x}_i,\ \delta_i,\ t_i)$，$i = 1, \ldots, N$。$\mathbf{x}_i$ 是表征第 i个样本的d维特征向量。δ_i 是事件发生指示变量。若 $\delta_i = 1$，则该个体在时间 t_i经历了该事件。反之，若 $\delta_i = 0$，则该个体未经历该事件，该样本为删失样本。 $t_i = \min\{t_i^c, t_i^e\}$是第 i个个体的删失时间 t_i^c 与实际事件发生时间 t_i^e 中的较小值。

生存模型分为三种：非参数、半参数和参数模型。非参数模型不对事件时间的基础分布作任何假设。这类模型主要用于数据可视化，因为它们编码了生存数据的汇总统计量。非参数生存模型包括 Kaplan-Meier [22], Nelson-Aalen [23], [24],和生命表 [25]。

半参数模型侧重于对事件随时间发生的瞬时风险进行建模。这类模型预测的是风险函数，而非生存函数

$$h(t|\mathbf{x}) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t | T \ge t, \mathbf{x})}{\Delta t}$$

或累积风险函数 $H(t|\mathbf{x}) = \int_0^t h(\tau|\mathbf{x})\, d\tau$。不过，仍然可以通过以下方式得到半参数模型的生存函数： $S(t|\mathbf{x}) = e^{-H(t|\mathbf{x})}$。风险函数有助于半参数模型将与总体相关、随时间变化的风险 $h_0(t)$（称为基线风险）与针对每个个体、随时间不变的风险 $h(\mathbf{x})$（称为风险函数）解耦。半参数生存模型中最著名的例子是 Cox比例风险（Cox比例风险模型） [26]。Cox比例风险模型将风险函数估计为

$$h(t|\mathbf{x}) = h_0(t) e^{\langle \beta, \mathbf{x} \rangle},$$

其中， $h_0(t)$ 是基线风险，风险函数为$h(\mathbf{x}) = e^{\langle \beta, \mathbf{x} \rangle}$。 $h_0(t)$ 可通过 Nelson-Aalen 估计量 [23], [24] 进行评估。CoxPH模型基于比例风险假设，即两个受试者的风险比随时间保持恒定。DeepSurv [27] 扩展了CoxPH模型，允许输入特征与输出风险之间存在非线性依赖关系。在 DeepSurv中，风险函数为 $h(\mathbf{x}) = \phi_w(\mathbf{x})$，其中 ϕ_w 是一个参数为 w的单输出神经网络。半参数模型的参数可以通过偏对数似然损失函数进行优化，该函数仅使用风险比即可评估。

参数模型假设事件的发生服从某种参数分布。与半参数模型相比，这类模型通常具有更强的表达能力，因为它们不受比例风险假设的限制，而这一假设在大多数现实场景中可能并不成立。DeepHit [28] 是一种基于时间离散化的参数模型，使用具有 sigmoid 激活函数的神经网络直接对生存时间进行建模。N-MTLR [29]则在多任务逻辑回归（MTLR）生存模型 [30] 的基础上，建立输入特征与预测概率之间的非线性依赖关系。

表 I 生存分析的机器学习模型。

模型	线性	比例风险	可微连续	
Cox比例风险模型 [26]	✓	✓	部分	✓
DeepSurv [27]	×	✓	部分	✓
DeepHit [28]	×	×	✓	×
N-MTLR [29]	×	×	✓	×
Nnet-Survival [31]	×	×	✓	×
PC-Hazard [32]	×	×	✓	✓
RSF [20]	×	×	×	✓

Nnet-Survival [31], [32], 也称为 Logistic Hazard, 使用神经网络对离散风险函数进行参数化。PC-Hazard [32], [33] 基于分段常数风险函数，因此得到的生存函数是连续的分段指数函数。PC-Hazard 将生存问题转化为泊松回归问题，使回归模型能够解决生存任务。随机生存森林（RSF） [20]通过递归构建一组二叉生存树来估计累积风险函数。与 [34],中一样，二叉树通过自助法构建。在每个节点处，从一组 $d' \leq d$ 特征中选出能使子节点间风险差异最大的特征。树的叶节点中的风险函数取平均值，作为集成风险函数。

表 I 总结了所述模型的特征。关于这些模型在联邦式设置下的比较，将在第 IV 节中详细讨论。

B. 联邦学习

联邦学习（FL） [5], [6] 是一种机器学习场景，其中一组客户端在中央服务器的协调下协作训练模型。在 FL 中，数据始终留在收集数据的设备上。典型的联邦式场景由 K 个客户端组成，每个客户端都持有一个私有数据集 D_k, $k = 1, \ldots, K$。给定一个参数为 w 的机器学习模型，FL 算法的目标是最小化关于 w 的全局损失函数 $\mathcal{L}$ ，即，

$$\min_w \mathcal{L}(w) = \min_w \sum_{k=1}^{K} \lambda_k \mathcal{L}_k(w).$$

此处， $\mathcal{L}_k$ 是由客户端 k 使用其私有数据 D_k计算得到的损失函数。 λ_k 是一组参数，用于衡量每个客户端对全局损失的贡献。通常，贡献与样本数量 $|D_k|$ 成正比，以促进在更多数据上评估的低方差损失。

联邦学习文献中，为优化 $\mathcal{L}(w)$ 提出的首个算法是联邦平均（FedAvg） [35]。FedAvg 会交替执行广播迭代和聚合迭代，直到收敛。在广播迭代期间，服务器会选择一部分 $K' \leq K$ 客户端，并向其发送当前模型参数 w。随后，这些客户端使用本地数据 D_k 对模型进行优化，并将更新后的参数发回服务器。通常，只需少量训练轮次即可完成本地训练。最后，服务器通过对客户端发来的更新取平均值来更新全局模型参数。

FedAvg 在模拟环境中表现稳定，但现实场景带来了若干挑战，限制了其泛化能力和收敛速度。具体而言，现实场景在计算能力和数据分布方面都存在异质性 [5]。为此，一些研究扩展了 FedAvg，以应对存在不可靠通信信道以及客户端之间数据非独立同分布的联邦 [36]–[41]。

为了测试联邦算法在异质联邦中的有效性，人们制定了通用的基准测试实践。用于联邦学习的最广泛异构数据集集合是 LEAF [42]。该库包含多个用于标准机器学习任务的异构数据集，例如图像分类和下一字符预测。S GDE [43] 提供了一个框架，用于从在联邦客户端上训练的隐私保护数据生成器构建合成数据集。借助 SGDE，每个生成器都体现了相应客户端的特征，从而生成固有的异构数据集。最后，其他研究 [44], [45] 探讨了基于狄利克雷分布的数据划分技术。这些技术将非联邦分类数据集中的数据样本分配给联邦中的各个客户端，并可控制异质性水平。

C. 联邦生存分析

联邦生存分析是指将生存模型整合到联邦算法中。大多数研究 [3], [7]–[16] 聚焦于 CoxPH模型的联邦式适配。具体而言，Cox部分对数似然损失函数在评估时需要访问整个数据集。不可分离性 给联邦应用带来了重大挑战。一些研究 [3], [16],提出了标准部分对数似然的替代形式，这些形式依赖于 [32]提出的生存模型离散化方法。另一些研究则探索了对替代似然进行一次性评估 [8]，或对牛顿-拉夫森优化方法进行分布式适配 [7], [10]。

跳出Cox模型的范畴， [46], 中的作者提出了一种用于联邦基因组学分析的安全Kaplan-Meier估计程序。其他研究 [17]–[19]聚焦于通用参数模型，包括深度神经网络。例如， [19],将生存问题表述为回归问题，并使用伪值作为目标标签。另一方面， [17], 通过弱监督注意力模块评估离散生存率。

隐私是分布式医疗保健应用的关键方面，已有多项研究采用差分隐私 [47] 来保护生存模型免受推断攻击 [17], [18]。具体而言， [18] 的作者通过添加后处理步骤来调节参数更新，并提升差分隐私模型的收敛，从而应对隐私级别与模型性能之间的权衡。其他研究 [12], [48] 依靠安全多方计算来防止分布式网络中的数据泄露，而同态加密 [46]

以及结合降维的自助法 [15] 则是较不常用的隐私保护方法。

关于目标应用，若干研究聚焦于基于癌症基因组图谱（TCGA）项目开展癌症研究中的基因组分析 [3], [17], [49], 。FLamby [49]提供了一套适用于医疗保健场景的联邦数据集，其中包括一个用于乳腺癌分析的生存数据集，名为 Fed-TCGA-BRCA。监测、流行病学和最终结果（SEER）数据库也已用于多项研究 [10], [13]。其他研究则探讨了中风检测 [8]、喉癌 [11], 以及 COVID-19 生存率 [14]。最后，关于联邦生存分析场景中的性能评估和结果可比性， [50] 提供了多种技术，可将现有生存数据集划分给联邦中的各个客户端，并对异质性水平进行控制。

据我们所知，FedSurF 是首个用于联邦生存分析的集成学习技术，它提供了一种不依赖比例假设的参数模型，并且只需一轮通信即可完成。

III. 联邦生存森林

在本节中，我们介绍将随机生存森林（RSF）算法调整为联邦学习设置的方法，并将其称为联邦生存森林（FedSurF）。我们的方法受到先前联邦集成学习研究的启发 [51], [52], ，这些研究通过整合各客户端本地集成中的基模型，在中央服务器上构建集成模型。具体而言，FedSurF通过汇总各客户端本地随机生存森林模型中表现最佳的树，在中央服务器上构建RSF。为此，我们评估了两种树选择技术：均匀概率选择方法，以及旨在选出表现最佳树的基于指标的方法。

FedSurF算法分为三个阶段：本地训练、树分配和树采样。在本地训练阶段，每个客户端 k 使用其本地数据集 D_k独立运行RSF算法，从而生成一个本地模型 M_k。每个 M_k 都由一组生存树 $\{T_1, \ldots, T_{N_k}\}$组成。为了针对 D_k优化 M_k的性能，可以通过在本地数据集 D_k上进行训练验证集划分并开展交叉验证，对每个 M_k 的超参数进行微调。这样便可选出本地树数量 N_k 以及树构建参数（最大树深度、每次分裂的最小样本数、每个叶节点的最小样本数和最大叶节点数）的最优取值。

每个客户端都训练好本地 RSF 模型后，树分配阶段由中央服务器确定每个客户端必须传输的树的数量，以达到所需的 N_S 棵树总数。为便于此过程，服务器必须获知每个本地模型中的树数量 N_k M_k。与包含模型参数的消息相比，此消息对通信信道的影响微乎其微，因为它只包含一个整数。收到各客户端可用树的数量 N_k 后，服务器

算法1： 联邦生存森林（FedSurF）

输入： K 个客户端， D_k， N_S，采样策略
输出： 生存树集成 M_S
初始化（服务器）： $N'_k \leftarrow 0$
初始化（客户端 k）： $M_k \leftarrow \emptyset$， $M'_k \leftarrow \emptyset$
对于 $k = 1$ 到 K 并行执行**则**
    $M_k \leftarrow$随机生存森林(D_k)
    $N_k \leftarrow |M_k|$
结束循环
循环 $i = 1$ 到 N_S **执行**
    $k \leftarrow$ 样本$(1, \{n\}_{n=1}^{K}, \{|D_k|\}_{k=1}^{K},$真$)$
    $N'_k \leftarrow N'_k + 1$
结束循环
循环 $k = 1$ 到 K 并行**执行**
    如果采样策略需要 IBS**则**
        $M'_k \leftarrow$样本$(N'_k,\ M_k,\ \{1/\text{IBS}_j\}_{j=1}^{N_k},$ False$)$
    else
        $M'_k \leftarrow$样本$(N'_k, M_k, \{1/N_k\}_{j=1}^{N_k},$否$)$
    结束 if
    发送 M'_k 到服务器
结束循环
$M_S \leftarrow \bigcup_{k=1}^{K} M'_k$
返回 M_S

执行 N_S 次迭代，每次按与数据集基数成正比的概率，为随机选定的客户端增加树计数器 $N'_k \leq N_k\,|D_k|$。这种非均匀概率有助于纳入数据集较大的客户端所训练的树，类似于 FedAvg [35] 的加权贡献。树分配阶段结束后，每个客户端都需要向服务器发送 N'_k 棵树，使服务器上的树总数等于 N_S。

在 FedSurF 的最后阶段，客户端需要从其本地模型中选择N'_k 棵树 M_k 并发送至中央服务器。选择过程可以采用均匀抽样策略，也可以根据验证指标为每棵树的选择概率赋予权重。为此，我们提出使用积分Brier评分（IBS）的倒数作为验证指标。IBS 是一种广泛用于衡量生存模型准确性的指标，详见第 IV-E 节。客户端可以计算每棵本地树的 IBS 分数，记为 IBS_j，其中 $j = 1, \ldots, N_k$，并以与 $1/IBS_j$ 成正比的概率选择 N'_k 棵树。该方法称为 FedSurF-IBS。如果客户端选择采用均匀抽样策略，则该方法简称为 FedSurF。最后，服务器通过聚合从各客户端接收的树来构建集成模型 M_S k。

FedSurF 的伪代码见算法1。函数 Sample(N, S, P, R) 用于从集合 S 中提取 N 个元素，每个元素的选取概率与 P 成正比。如果将 R 设置为 True，则采用有放回抽样。

表 I 实验中包含的生存数据集。

数据集	样本	删失	数值特征	类别特征
GBSG2 [53]	686	44%	5	3
METABRIC [27]	1904	58%	5	3
AIDS [54]	2839	62%	1	3
FLCHAIN [55]	7874	28%	6	4
SUPPORT [56]	9105	68%	24	11

IV. 实验

本节通过一系列全面的实验，实证展示 FedSurF 在解决具有较强泛化能力且能适应非独立同分布数据的生存问题方面的有效性。源代码可在 https://github.com/archettialberto/federated_survival_forests 获取。

A. 数据集

本研究在五个生存数据集上开展实验：德国乳腺癌研究组2（GBSG2） [53], 乳腺癌分子分类国际联盟（METABRIC） [27],澳大利亚艾滋病生存数据集（AIDS） [54], 血清游离轻链检测数据集（FLCHAIN） [55], 以及治疗预后、偏好、结局与风险理解研究（SUPPORT） [56]。这些数据集的相关汇总统计量见表II。

B. 联邦模拟技术

在实验过程中，我们首先将一个非联邦生存数据集 D 划分为训练数据集和测试数据集，前者占总样本数的80%，后者占20%。测试数据集不再进行进一步修改，仅用于对每个模型进行最终评估。随后，我们将训练数据集中的每个样本分配给一个联邦中的某个客户端。每个联邦由10个客户端组成（$K = 10$）。为确定每个样本分配到哪个客户端，我们采用两种技术。第一种技术以相同的概率将训练数据集中的每个样本分配给一个可用客户端。这种技术使联邦中各客户端的数据分布均匀，从而促进联邦算法的收敛。图1第一行展示了采用均匀分配技术时客户端数据的Kaplan-Meier估计量$\hat{}$ $S_k(t)$。

第二种划分技术遵循 [50]中所述的标签偏斜划分算法。在这种情况下，将数据样本分配给联邦客户端，使所得的标签分布具有异质性。该技术基于狄利克雷分布，灵感来自分类联邦数据集中采用的类似做法 [44], [45]。其目标是模拟数据分布不均匀的真实联邦环境。为此，我们应用 [50]中的标签偏斜划分算法 $K = 10$ 和 $\alpha = 8.0$，并规定每个数据集至少包含 25 个样本。图 1 的第二行展示了标签偏斜划分下每个客户端数据集 D_k的 Kaplan-Meier估计量 $\hat{} S_k(t)$。

在数据分配程序的基础上，我们对每个客户端数据集采用80%-20%划分，以提取本地验证集划分，用于参数调优和提前停止。最后，假设每个客户端始终可用，且通信信道完全可靠。

C. 基线模型

在实验评估中，我们将所提出的 FedSurF 和 FedSurF-IBS 算法的性能与第 II-A 节介绍的最先进的生存模型进行比较：Cox比例风险（Cox比例风险模型） [26], DeepSurv [27],DeepHit [28], 神经多任务逻辑回归（N-MTLR） [29], Nnet-Survival [31], 以及分段常数风险（PC-Hazard） [32]。FedSurF 和 FedSurF-IBS 的超参数通过交叉验证进行优化。对于 Cox比例风险模型，参数数量等于输入空间中的特征数。DeepSurv、DeepHit、N-MTLR、Nnet-Survival 和 PC-Hazard 的神经网络架构均由两个各含 32 个神经元的全连接层组成，随后接入 ReLU激活函数和一个 Dropout层，以防止过拟合。DeepSurv 的最后一个全连接层输出一个符合比例风险假设的标量，而其他模型则输出 10 个值，每个值对应一个离散化时刻。

D. 训练

在实验中，我们分别在本地设置和联邦设置下训练了每个模型。在本地设置中，客户端仅使用本地数据训练和验证模型。本地评估有助于了解参与联邦学习算法是否对客户端有益。

RSF 的实现基于 scikit-survival Python库 [57],, 而其他可微模型则依赖 PyCox [58]。RSF 的超参数调优通过交叉验证进行。可微模型使用 Adam优化器进行训练, 学习率为 0.01，最多训练 500 个训练轮次，并采用提前停止。对于 FedSurF 和 FedSurF- IBS，联邦训练按照第 III 节所述的算法进行。相比之下，FedAvg [35]应用于其他所有模型。对于比例风险模型（CoxPH 和 DeepSurv），我们假设每个客户端都能访问相同的全局基线风险 Nelson-Aalen估计。为解决异构联邦中的收敛问题，我们运行了 500 轮 FedAvg，每轮进行一个本地训练轮次，并选择验证指标最佳的参数，在测试集上进行最终评估。

E. 评估

在我们的评估中，我们使用一致性指数（C指数）和积分 Brier评分（IBS）来评估模型在测试集上的表现。C指数 [59]衡量可比较样本中预测结果与真实结果一致的比例，其中

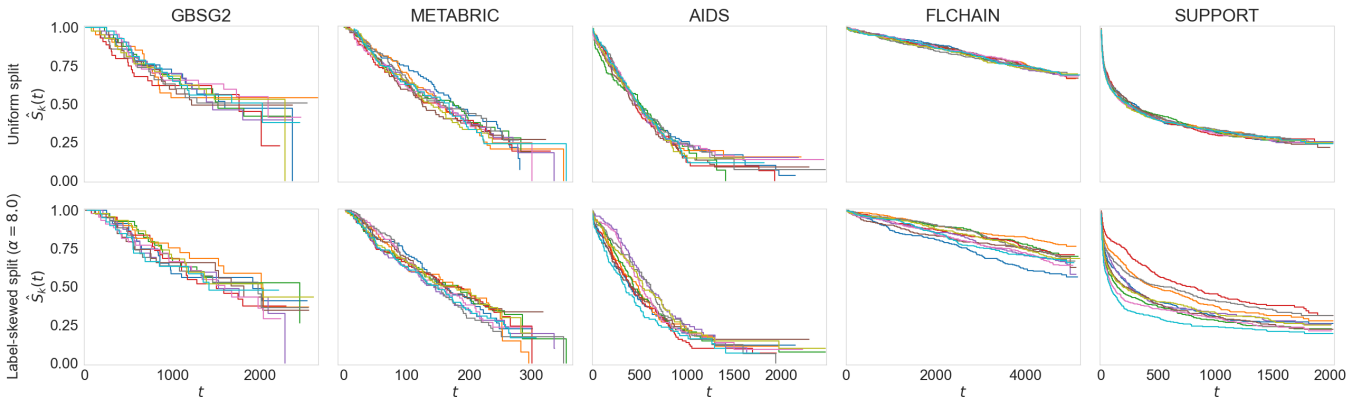


图 1. 均匀数据分配（第一行）和标签偏斜分配 [50] （第二行）下，每个客户端数据集 D_k 的 Kaplan-Meier估计量 $\hat{S}_k(t)$，其中标签偏斜分配中 $\alpha = 8.0$ 。每条曲线对应一个客户端 k。

预测结果与真实结果一致。如果两个样本中至少有一个未被删失，则认为该样本对具有可比性。C指数反映了模型的区分能力，表示模型根据实际时间结果正确排序的样本比例。随机猜测模型的C指数为0.5，而具有完美预测能力的模型的C指数为1.0。

Brier分数 [21]是一种校准指标，用于衡量模型在给定时间预测的生存概率与真实生存状态之间的平方差。IBS 对 Brier分数随时间进行积分，将模型的校准水平归纳为单一数值。随机猜测模型的 IBS 为 0.25，而具有完美知识的模型的 IBS 为 0.0。因此，IBS 越低，模型越好。

为考虑删失分布，我们在指标估计过程中应用了删失逆概率加权（IPCW） [59], [60]，以确保结果无偏。

F. 结果

表III和表IV总结了采用均匀数据划分构建的联邦中各生存模型的性能，表V和表VI则总结了采用标签偏斜划分构建的联邦中的性能。下面，我们将对所得结果进行详细分析。

客户端加入联邦学习过程是否有益？ 所有结果一致表明，平均而言，最佳性能始终出现在联邦式列中。这说明，对于给定的生存模型，加入联邦通常有利于提升判别能力和校准能力。

当数据均匀分布时，*FedSurF* 是否优于其他方法？为回答这一问题，我们参考了表III和表IV。从结果可以看出，FedSurF 并未始终优于其他最先进的模型。不过，无论在区分能力还是校准方面，其表现都与其他模型相当。进一步观察发现，FedSurF 在较小的数据集（GBSG2 和 METABRIC）上表现尤为出色；而在较大的数据集（AIDS、FLCHAIN 和 SUPPORT）上，其结果与深度学习模型相当。这表明，当数据均匀分布时，基于集成的方法与基于神经网络的方法之间的模型选择并不是特别重要。不过，我们强调，FedSurF 和 FedSurF-IBS 仅用一个通信轮次便取得了所报告的性能；在存在通信信道约束的联邦环境中，这可能是一项相关优势。

数据均匀分布时，*IBS*采样是否相关？当数据均匀分布时，FedSurF-IBS并未始终优于FedSurF。对比表III和表IV即可看出这一点。这一发现表明，在数据独立同分布的情形下，要实现最佳性能并不需要使用IBS采样。这一结果在意料之中，因为IBS采样尤其有助于从最佳客户端中选出最佳树。在独立同分布数据的情况下，各客户端的本地数据分布大致相同，因此每个客户端生成的树的质量应该相近。在这些条件下，IBS并不能帮助实现最佳性能。

当数据具有异构性时，*FedSurF* 的表现是否优于其他方法？对此，我们参考表V和表VI。结果表明，除 SUPPORT 外，FedSurF-IBS 在所有数据集上都获得了最佳或非常接近最佳的指标。具体而言，在 GBSG2、METABRIC 和 AIDS 上，它的 C指数均为最佳。此外，其 IBS 也非常低，并且处于略优模型的标准差范围内。根据这些结果可以得出结论：在考察 C指数和 IBS指标时，采用 IBS采样的 FedSurF 是异构联邦中表现良好的算法。它在不同的标签偏斜数据集上展现出稳健的性能，相较于其他模型始终有所提升。

当数据具有异构性时，*IBS*采样是否具有相关性？表V和表VI所呈现的异构数据集结果表明，对于除

表III 均匀划分的数据集上评估的生存模型的 C指数。为便于阅读，每项指标均乘以 100。我们报告 5 次运行的均值和标准差。最佳结果(↑)以粗体突出显示。

数据集	GBSG2		METABRIC		AIDS		FLCHAIN		SUPPORT	
模型	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式
Cox比例风险模型	61.2±4.1	68.7±8.5	58.8±1.9	62.0±1.7	51.9±0.7	53.8±1.6	92.3±1.6	93.4±0.4	79.9±0.2	82.5±0.3
DeepSurv	60.7±4.0	69.3±7.5	59.7±1.7	**63.0±2.7**	52.5±0.2	53.7±1.3	93.5±0.4	94.2±0.2	79.9±0.4	83.1±0.4
DeepHit	59.8±5.3	68.5±7.5	57.3±1.3	61.1±3.0	52.6±0.5	53.8±1.6	93.1±0.3	93.8±0.4	81.7±0.4	83.5±0.6
N-MTLR	61.5±4.4	69.2±7.1	58.3±1.6	62.2±2.5	53.1±0.8	**54.2±1.6**	93.6±0.3	**94.2±0.3**	80.6±0.4	83.0±0.6
Nnet-Survival	60.5±7.9	66.7±7.7	49.4±3.9	60.4±4.1	50.8±1.2	54.0±1.6	90.7±1.3	94.1±0.4	80.9±0.3	**83.8±0.6**
PC-Hazard	60.3±8.2	67.3±7.4	50.2±3.9	59.8±3.8	51.0±1.3	53.3±1.8	89.8±0.8	94.1±0.3	80.7±0.3	81.9±1.2
FedSurF	66.5±5.7	**72.3±4.7**	59.9±2.1	62.2±1.9	52.4±0.6	54.1±1.3	93.4±0.3	93.5±0.3	80.4±0.4	81.0±0.4
FedSurF-IBS	–	71.8±5.4	–	62.4±2.0	–	53.8±1.3	–	93.5±0.3	–	80.7±0.6

表IV 在均匀划分的数据集上评估的生存模型的 IBS。为便于阅读，每项指标均乘以 100。我们报告 5 次运行的均值和标准差。最佳结果（↓）以粗体突出显示。

数据集	GBSG2		METABRIC		AIDS		FLCHAIN		SUPPORT	
模型	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式
Cox比例风险模型	20.2±1.1	18.5±1.6	17.9±1.9	17.4±2.3	15.1±1.5	14.7±1.5	5.6±0.4	5.4±0.2	13.3±0.2	12.7±0.1
DeepSurv	21.2±1.2	**17.9±1.2**	17.8±1.5	**16.6±1.5**	15.2±1.2	14.7±1.5	5.6±0.6	4.2±0.1	13.9±0.2	12.2±0.2
DeepHit	21.8±2.0	19.5±1.1	18.7±1.7	17.6±1.6	16.0±1.4	14.8±1.5	5.8±0.2	4.5±0.1	13.7±0.6	12.7±0.4
N-MTLR	21.3±1.6	18.3±1.0	18.1±1.5	17.1±1.7	14.9±1.4	14.7±1.4	4.8±0.1	**4.2±0.1**	14.5±0.3	**11.9±0.4**
Nnet-Survival	23.0±2.1	19.5±1.0	23.9±2.7	17.7±1.6	15.4±1.5	15.0±1.5	6.5±0.4	4.5±0.2	12.9±0.3	12.3±0.4
PC-Hazard	22.4±1.9	19.6±0.8	23.7±2.4	18.0±1.6	15.4±1.6	15.1±1.5	6.8±0.4	4.5±0.2	13.0±0.3	12.4±0.4
FedSurF	19.3±1.2	18.5±1.2	17.6±1.3	16.8±1.2	14.8±1.5	14.7±1.5	4.6±0.1	4.5±0.1	18.2±0.5	18.1±0.5
FedSurF-IBS	–	18.4±1.2	–	16.7±1.3	–	**14.7±1.5**	–	4.4±0.1	–	15.9±0.5

表V：在标签偏斜数据集上评估的生存模型的C指数。为便于阅读，每项指标均乘以100。我们报告5次运行的均值和标准差。最佳结果（↑）以粗体突出显示。

数据集	GBSG2		METABRIC		AIDS		FLCHAIN		SUPPORT	
模型	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式
Cox比例风险模型	60.3±2.8	68.4±7.8	57.3±0.9	60.6±3.8	52.2±1.4	54.2±1.1	91.5±2.8	77.6±22.5	80.0±0.2	82.3±0.3
DeepSurv	59.8±2.2	67.5±8.5	58.7±1.8	62.1±3.7	53.0±0.6	54.5±1.6	92.3±3.1	81.9±18.6	80.0±0.5	80.7±2.2
DeepHit	60.9±5.3	68.5±7.0	56.2±1.0	60.4±2.9	53.2±1.1	53.9±0.9	93.2±0.3	93.7±0.1	81.0±0.2	82.9±0.6
N-MTLR	61.8±8.9	69.3±7.6	59.0±1.9	61.7±3.3	53.0±0.7	53.6±2.1	93.6±0.2	**94.0±0.2**	80.3±0.3	81.1±0.4
Nnet-Survival	59.5±8.5	68.2±7.7	49.1±3.6	59.5±3.7	51.3±1.5	54.2±2.3	90.6±1.2	93.9±0.3	80.7±0.5	**83.5±0.7**
PC-Hazard	60.9±8.0	68.9±7.3	49.4±3.8	59.2±3.6	50.9±0.6	53.6±2.6	89.9±1.1	93.9±0.3	80.6±0.5	80.6±0.8
FedSurF	66.8±5.4	71.9±5.0	59.9±1.4	61.8±2.9	53.2±0.8	55.0±0.9	93.4±0.3	93.5±0.3	80.3±0.4	81.3±0.4
FedSurF-IBS	–	**72.4±5.4**	–	**62.1±2.4**	–	**55.1±0.6**	–	93.5±0.2	–	80.6±0.7

表VI 标签偏斜数据集上评估的生存模型 IBS。为便于阅读，每项指标均乘以100。我们报告5次运行的均值和标准差。最佳结果（↓）以粗体突出显示。

数据集	GBSG2		METABRIC		AIDS		FLCHAIN		SUPPORT	
模型	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式	本地	联邦式
Cox比例风险模型	20.1±1.2	18.5±1.4	18.3±1.6	17.3±2.3	15.1±1.5	14.7±1.5	5.7±0.6	8.5±4.0	13.5±0.2	12.8±0.2
DeepSurv	20.6±1.5	**18.2±1.0**	18.7±1.6	**16.8±1.7**	15.2±1.1	14.7±1.5	5.5±0.6	9.0±4.4	14.0±0.3	12.6±0.5
DeepHit	21.7±2.0	19.2±0.7	18.6±1.7	17.7±1.5	16.0±1.3	14.7±1.5	5.7±0.2	4.5±0.1	14.4±0.6	12.8±0.4
N-MTLR	20.8±0.9	18.7±1.5	18.2±1.7	17.4±1.9	15.0±1.4	14.7±1.4	4.8±0.1	4.5±0.2	14.7±0.3	13.6±0.7
Nnet-Survival	23.6±2.5	19.1±1.5	23.9±2.4	17.9±1.4	15.4±1.6	14.9±1.6	6.6±0.5	4.5±0.2	13.0±0.4	**12.6±0.6**
PC-Hazard	22.6±2.1	19.0±1.5	23.9±2.5	18.2±1.6	15.4±1.5	15.0±1.6	6.8±0.4	4.5±0.2	13.1±0.4	12.7±0.5
FedSurF	19.3±1.2	18.5±1.3	17.8±1.2	16.9±1.3	14.9±1.5	14.7±1.5	4.7±0.1	4.5±0.1	18.5±0.6	18.0±0.6
FedSurF-IBS	–	18.4±1.1	–	16.9±1.2	–	**14.7±1.5**	–	**4.4±0.1**	–	15.8±0.5

在 SUPPORT 上，FedSurF-IBS 的表现优于 FedSurF。这表明，与数据均匀划分的联邦相比，在数据具有异质性的联邦中，IBS 采样具有重要意义且能带来收益。事实上，它能够从最佳客户端中选出最佳树，从而提升集成模型的整体性能。这凸显了在为 FedSurF 选择合适的采样方法时，考虑数据分布特征的重要性。关于 SUPPORT 数据集的结果，尽管进行了广泛的超参数搜索，在由 10 个客户端组成的联邦中，生存森林的 Brier 分数仍高于神经模型。我们认为，较低的 IBS 归因于模型的树结构；与基于神经网络的方法相比，这类模型对 SUPPORT 特定生存率的建模可能不够理想。本地结果证实了这一判断：即使不考虑联邦学习，树模型的平均表现也低于神经模型。尽管如此，在联邦式实验中应用 IBS 采样可将 IBS 降低 0.02 以上，相较于均匀采样取得了显著提升。

五、结论

本文介绍了联邦生存森林（FedSurF），这是将随机生存森林（RSF）算法扩展到联邦学习场景的成果。我们提出的算法基于从联邦中的最佳客户端采样最佳本地树，并在服务器上构建 RSF。我们在两种不同的联邦场景中进行了广泛实验，以验证 FedSurF 的有效性：一种是数据均匀划分，另一种是包含异构数据。结果表明，当数据均匀分布时，FedSurF 的表现与最先进模型相当；而在异构联邦中，其性能优势则十分明显。此外，FedSurF 仅依赖客户端与服务器之间的一轮通信，而非多次迭代更新，因此在带宽要求严格的联邦中是一种高效且实用的解决方案。

致谢

本项目得到 AI-SPRINT：安全隐私保护计算连续体中的人工智能（欧盟 H2020 资助协议编号 101016577）以及 FAIR：未来人工智能研究（下一代欧盟、PNRR-PE-AI计划、M4C2、投资1.3、人工智能专项）的支持。

参考文献

[1] J. P. Klein 和 M. L. Moeschberger，《生存分析：删失数据和截断数据的技术》。Springer，2003，卷 1230。[2] P. Wang、Y. Li 和 C. K. Reddy，“用于生存分析的机器学习：综述”，《ACM Computing Surveys (CSUR)》，第 51 卷，36，2019 年 6 月，1no. pp. – 。[3] M. Andreux、A. Manoel、R. Menuet、C. Saillard 和 C. Simpson，“基于离散时间 Cox 模型的联邦生存分析”，《arXiv 预印本 arXiv:2006.08997》，2020 年。[4] N. Rieke、J. Hancox、W. Li、F. Milletari、H. R. Roth、S. Albarqouni、S. Bakas、M. N. Galtier、B. A. Landman、K. Maier-Hein 等，“联邦学习下数字健康的未来”，《NPJ 数字医学》，第 3 卷，第 1 期，第 1–7 页，2020 年。

[5] T. Li, A. K. Sahu, A. Talwalkar, 和 V. Smith，《联邦学习：挑战、方法与未来方向》，IEEE信号处理杂志，第 37 卷，第 3 期，第 50–60 页，2020 年。[6] P. Kairouz, H. B. McMahan, B. Avent, A. Bellet, M. Bennis, A. N. Bhagoji, K. Bonawitz, Z. Charles, G. Cormode, R. Cummings 等，《联邦学习的进展与开放问题》，机器学习基础与趋势，第 14 卷，第 1–2 期，第 1–210 页，2021 年。[7] C.-L. Lu, S. Wang, Z. Ji, Y. Wu, L. Xiong, X. Jiang, 和 L. Ohno- Machado，《Webdisco：一种无需共享患者级数据的分布式 Cox 模型学习 Web 服务》，美国医学信息学会杂志，第 22 卷，第 6 期，第 1212–1219 页，2015 年。[8] R. Duan, C. Luo, M. J. Schuemie, J. Tong, C. J. Liang, H. H. Chang, M. R. Boland, J. Bian, H. Xu, J. H. Holmes, C. B. Forrest, S. C. Morton, J. A. Berlin, J. H. Moore, K. B. Mahoney, 和 Y. Chen，《从局部到全局学习：一种用于生存时间数据建模的高效分布式算法》，美国医学信息学会杂志，第 27 卷，第 7 期，第 1028–1036 页，2020 年 7 月。[ ] 10 1093 044 在线。可访问：[9] https://doi.org/ . /jamia/ocaa S. Banerjee, G. N. Sofack, T. Papakonstantinou, D. Avraam, P. Burton, D. Z¨oller, 和 T. R. P. Bishop，《dsSurvival：用于 DataSHIELD 联邦个体患者荟萃分析的隐私保护生存模型》，BMC研究笔记，第 15 卷，第 1 期，第 197 页，2022 年 12 月。 [在线]。可访问：https://bmcresnotes.biomedcentral.com/articles/10 13 104 1186 022 06085 1 . /s - - -[10]W. Dai, X. Jiang, L. Bonomi, Y. Li, H. Xiong, 和 L. Ohno-Machado，《VERTICOX：使用交替方向乘子法的垂直分布式 Cox 比例风险模型》，IEEE知识与数据工程汇刊，第 34 卷，第 2 期，第 996–1010 页，2022 年 2 月。 [在线]。可访问：https: //ieeexplore.ieee.org/document/9076318/ [11] C. R. Hansen, G. Price, M. Field, N. Sarup, R. Zukauskaite, J. Johansen, J. G. Eriksen, F. Aly, A. McPartlin, L. Holloway, D. Thwaites, 和 C. Brink，《通过开源联邦学习开发的喉癌生存模型》，放射治疗与肿瘤学，第 176 卷，第 179–186 页，2022 年 11 月。 [在线]。可访问：https: // linkinghub.elsevier.com/retrieve/pii/S 0167814022044930 [12] B. Kamphorst, T. Rooijakkers, T. Veugen, M. Cellamare, 和 D. Knoors，《在保护隐私的同时，基于垂直分区数据准确训练 Cox 比例风险模型》，BMC医学信息学与决策制定，第 22 卷，第 1 期，第 49 页，2022 年 12 月。[在线]。可访问：https://bmcmedinformdecismak.biomedcentral.com / articles/10.1186/s12911-022-01771-3[13] C. Masciocchi, B. Gottardelli, M. Savino, L. Boldrini, A. Martino, C. Mazzarella, M. Massaccesi, V. Valentini, 和 A. Damiani，《联邦 Cox 比例风险模型：面向个性化医疗的多中心隐私保护 LASSO 特征选择与生存分析》，载于 2022 年 IEEE第35届基于计算机的医疗系统国际研讨会（CBMS）。中国深圳：IEEE，2022 年 7 月，第 25–31 页。 [在线]。可访问：https://ieeexplore.ieee.org/document/9867090/ [14] X. Wang, H. G. Zhang, X. Xiong, C. Hong, G. M. Weber, G. A. Brat, C.- L. Bonzel, Y. Luo, R. Duan, N. P. Palmer 等，《Survmaximin：一种用于迁移生存风险预测模型的稳健联邦方法》，生物医学信息学杂志，第 134 卷，第 104176 页，2022 年。[15] A. Imakura, R. Tsunoda, R. Kagawa, K. Yamagata, 和 T. Sakurai，《DC- COX：用于多方隐私保护生存分析的数据协作 Cox 比例风险模型》，生物医学信息学杂志，第 137 卷，第 104264 页，2023 年 1 月。 [在线]。可访问：https://linkinghub.elsevier.com/retrieve/pii/S 1532046422002696 [16] D. K. Zhang, F. Toni, 和 M. Williams，《一种具有非比例风险的联邦 Cox 模型》，载于《医疗保健中的多模态人工智能》。Springer，2023 年，第 171–185 页。[17] M. Y. Lu, R. J. Chen, D. Kong, J. Lipkova, R. Singh, D. F. Williamson, T. Y. Chen, 和 F. Mahmood，《面向十亿像素全视野图像计算病理学的联邦学习》，医学图像分析，第 76 卷，第 102298 页，2022 年 2 月。 [在线]。可访问：https://linkinghub.elsevier.com/retrieve/pii/S 1361841521003431 [18] S. Rahimian, R. Kerkouche, I. Kurth, 和 M. Fritz，《医疗数据差分隐私联邦生存分析中的实践挑战》，载于健康、推断与学习会议。PMLR，2022 年，第 411–425 页。[19] M. M. Rahman 和 S. Purushotham，《Fedpseudo：基于伪值的联邦生存分析深度学习模型》，arXiv预印本 arXiv:2207.05247，2022 年。

[20] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, 和 M. S. Lauer, 《随机生存森林》, 应用统计年鉴, 第 2 卷, 第 3 期, 第 841–860 页, 2008 年。[21] E. Graf, C. Schmoor, W. Sauerbrei, 和 M. Schumacher, 《生存数据预后分类方案的评估与比较》, 医学统计学, 第 18 卷, 第 17-18 期, 第 2529–2545 页, 1999 年。[22] E. L. Kaplan 和 P. Meier, 《基于不完整观测的非参数估计》, 美国统计学会杂志, 第 53 卷, 第 282 期, 第 457–481 页, 1958 年。[23] W. Nelson, 《删失失效数据风险绘图的理论与应用》, Technometrics, 第 14 卷, 第 4 期, 第 945–966 页, 1972 年。[24] O. Aalen, 《一类计数过程的非参数推断》, 统计年鉴, 第 701–726 页, 1978 年。[25] S. J. Cutler 和 F. Ederer, 《在生存分析中最大限度地利用生命表法》, 慢性疾病杂志, 第 8 卷, 第 6 期, 第 699–712 页, 1958 年。[26] D. R. Cox, 《回归模型与生命表》, 英国皇家统计学会杂志 B辑（方法论）, 第 34 卷, 第 2 期, 第 187 220, 1972 [ ] 2985181 – 页。[27] 在线。可访问：http://www.jstor.org/stable/ J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, 和 Y. Kluger, 《DeepSurv：使用 Cox 比例风险深度神经网络的个性化治疗推荐系统》, BMC医学研究方法学, 第 18 卷, 第 1 期, 第 1–12 页, 2018 年。[28] C. Lee, W. Zame, J. Yoon, 和 M. Van Der Schaar, 《DeepHit：一种针对竞争风险的生存分析深度学习方法》, 载于 AAAI人工智能会议论文集, 第 32 卷, 第 1 期, 2018 年。[29] S. Fotso, 《基于多任务框架的生存分析深度神经网络》, arXiv预印本 arXiv:1801.05512, 2018 年。[30] C.-N. Yu, R. Greiner, H.-C. Lin, 和 V. Baracos, 《将患者特异性癌症生存分布作为一系列依赖回归量进行学习》, 神经信息处理系统进展, 第 24 卷, 2011 年。[31] M. F. Gensheimer 和 B. Narasimhan, 《一种适用于神经网络的可扩展离散时间生存模型》, PeerJ, 第 7 卷, 第 e6257 页, 2019 年。[32] H. Kvamme 和 Ø. Borgan, 《使用神经网络进行连续时间和离散时间生存预测》, 生存数据分析, 第 27 卷, 第 4 期, 第 710–736 页, 2021 年。[33] A. Bender, D. R¨uga mer, F. Scheipl, 和 B. Bischl, 《用于生存分析的通用机器学习框架》, 载于欧洲机器学习与数据库知识发现联合会议。Springer, 2021 年, 第 158–173 页。[34] L. Breiman, J. H. Friedman, R. A. Olshen, 和 C. J. Stone, 《分类回归树》。Routledge, 2017 年。[35] B. McMahan, E. Moore, D. Ramage, S. Hampson, 和 B. A. y Arcas, 《基于去中心化数据的深度网络高效通信学习》, 载于人工智能与统计学。PMLR, 2017 年, 第 1273–1282 页。[36] S. Reddi, Z. Charles, M. Zaheer, Z. Garrett, K. Rush, J. Kone˘cn`y, S. Kumar, 和 H. B. McMahan, 《自适应联邦优化》, arXiv预印本 arXiv:2003.00295, 2020 年。[37] J. Wang, Z. Charles, Z. Xu, G. Joshi, H. B. McMahan, M. Al-Shedivat, G. Andrew, S. Avestimehr, K. Daly, D. Data 等, 《联邦优化指南》, arXiv预印本 arXiv:2107.06917, 2021 年。[38] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, 和 V. Smith, 《异构网络中的联邦优化》, 机器学习与系统论文集, 第 2 卷, 第 429–450 页, 2020 年。[39] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, 和 A. T. Suresh, 《SCAFFOLD：用于联邦学习的随机控制平均》, 载于国际机器学习会议。PMLR, 2020 年, 第 5132–5143 页。[40] D. A. E. Acar, Y. Zhao, R. M. Navarro, M. Mattina, P. N. Whatmough, 和 V. Saligrama, 《基于动态正则化的联邦学习》, arXiv预印本 arXiv:2111.04263, 2021 年。[41] D. Caldarola, B. Caputo, 和 M. Ciccone, 《通过寻找平坦极小值提升联邦学习的泛化能力》, 载于欧洲计算机视觉会议。Springer, 2022 年, 第 654–672 页。[42] S. Caldas, S. M. K. Duddu, P. Wu, T. Li, J. Kone˘cn`y, H. B. McMahan, V. Smith, 和 A. Talwalkar, 《LEAF：联邦场景基准》, arXiv预印本 arXiv:1812.01097, 2018 年。[43] E. Lomurno, A. Archetti, L. Cazzella, S. Samele, L. Di Perna, 和 M. Matteucci, 《SGDE：面向跨数据孤岛联邦学习的安全生成式数据交换》, 载于 AIPR 2022, 国际人工智能与模式识别会议, 2022 年。[44] T.-M. H. Hsu、H. Qi 和 M. Brown, “衡量非相同数据分布对联邦视觉分类的影响”, arXiv 预印本 arXiv:1909.06335, 2019。[45] Q. Li、Y. Diao、Q. Chen 和 B. He, “非独立同分布数据孤岛上的联邦学习：一项实验研究”, 载于《2022年IEEE第38届国际数据工程会议（ICDE）》。IEEE, 2022年, 第965–978页。[46] D. Froelicher、J. R. Troncoso-Pastoriza、J. L. Raisaro、M. A. Cuendet、J. S. Sousa、H. Cho、B. Berger、J. Fellay 和 J.-P. Hubaux, “利用多方同态加密实现精准医学中真正保护隐私的联邦分析”, 《自然通讯》, 第12卷, 第1期, 文章号5910, 2021年10月。 [在线]。网址：https: // www.nature.com/articles/s41467-021-25972-y [47] C. Dwork, “差分隐私：研究成果综述”, 载于《计算模型理论与应用国际会议》。Springer, 2008年, 第1–19页。[48] T. Marchand、B. Muzellec、C. Beguier、J. O. d. Terrail 和 M. Andreux, “SecureFedYJ：一种用于联邦学习的安全特征高斯化协议”, 2022年10月, arXiv:2210.01639 [cs]。 [在线]。网址：http://arxiv.org/abs/ 2210 01639 。J. O. d. Terrail、S.-S. Ayed、E. Cyffers、[49] F. Grimberg、C. He、R. Loeb、P. Mangold、T. Marchand、O. Marfoq、E. Mushtaq、B. Muzellec、C. Philippenko、S. Silva、M. Tele´nczuk、S. Albarqouni、S. Avestimehr、A. Bellet、A. Dieuleveut、M. Jaggi、S. P. Karimireddy、M. Lorenzi、G. Neglia、M. Tommasi 和 M. Andreux, “FLamby：真实医疗保健场景下跨数据孤岛联邦学习的数据集与基准”, 2022年10月, arXiv:2210.04620 [cs]。 [在线]。网址：http://arxiv.org/abs/ 2210 04620 。A. Archetti、E. Lomurno、F. Lattari、[50] A. Martin 和 M. Matteucci, “用于联邦生存分析模拟的异构数据集”, 载于《2023年ACM/SPEC性能工程国际会议论文集》, ICPE '23 Companion。美国纽约：计算机协会, 2023年, 第173–180页。 [ ] 10 1145 3578245 3584935 在线。网址：[51] https://doi.org/ 。/ 。A.-C. Hauschild、M. Lemanczyk、J. Matschinske、T. Frisch、O. Zolotareva、A. Holzinger、J. Baumbach 和 D. Heider, “联邦随机森林可提升预测模型在多种医疗保健应用中的本地性能”, 《生物信息学》, 第38卷, 第8期, 第2278–2286页, 2022年。[52] M. Gencturk、A. A. Sinaci 和 N. K. Cicekli, “Bofrf：一种用于水平分区数据的新型提升式联邦随机森林算法 ”, 《IEEE Access》, 第10卷, 第89835–89851页, 2022年。[53] M. Schumacher、G. Bastert、H. Bojar、K. H¨ubner、M. Olschewski、W. Sauerbrei、C. Schmoor、C. Beyerle、R. Neumann 和 H. Rauschecker, “评估淋巴结阳性乳腺癌患者激素治疗及化疗疗程的随机2×2试验。德国乳腺癌研究组。”《临床肿瘤学杂志》, 第12卷, 第10期, 第2086–2093页, 1994年。[54] B. Ripley、B. Venables、D. M. Bates、K. Hornik、A. Gebhardt 和 D. Firth, “R软件包：Mass”, 2022年7月27日。 [在线]。网址：https://stat.ethz.ch/R-manual/R-devel/library/MASS/html/00Index.html [55] T. Therneau、T. Lumley、E. Atkinson 和 C. Crowson, “R软件包：survival”, 2023年1月9日。 [在线]。网址：https://stat.ethz.ch/R-manual/R-devel/library/survival/html/00Index.html[56] 范德比尔特大学生物统计学系, “范德比尔特生物统计学数据集”, 2022年12月1日。[在线]。网址：http://hbiostat.org/data [57] S. P¨olsterl, “scikit-survival：一个基于scikit-learn构建的事件发生时间分析库”, 《机器学习研究杂志》, 第21卷, 第212期, 第1–6页, 2020年。 [在线]。网址：http: // jmlr.org/papers/v21/20-729.html [58] H. Kvamme、Ø. Borgan 和 I. Scheel, “利用神经网络和Cox回归进行事件发生时间预测”, arXiv预印本 arXiv:1907 00825, 2019。。[59] H. Uno、T. Cai、M. J. Pencina、R. B. D'Agostino 和 L.-J. Wei, “关于使用删失生存数据评估风险预测方法总体适用性的C统计量”, 《医学统计学》, 第30卷, 第10期, 第1105–1117页, 2011年。[60] J. M. Robins 和 A. Rotnitzky, “利用替代标志物恢复信息并调整依赖性删失”, 载于《AIDS流行病学》。Springer, 1992年, 第297–331页。