

Personalized Federated Learning with Feature Alignment and Classifier Collaboration

Jian Xu, Xinyi Tong, Shao-Lun Huang

摘要

数据异构性是联邦学习中最具挑战性的问题之一，它激发了各种方法来为参与的客户学习个性化模型。基于深度神经网络的任务中的一种此类方法是采用共享特征表示并为每个客户端学习定制的分类器头。然而，以前的工作在局部表示学习中没有利用全局知识，也忽略了局部分类器头之间的细粒度协作，这限制了模型的泛化能力。在这项工作中，通过利用全局语义知识来学习更好的表示来进行显式的局部全局特征对齐。此外，将每个客户端的分类器组合的好处量化为组合权重的函数，并导出用于估计最佳权重的优化问题。最后，对各种异构数据场景的基准数据集的广泛评估结果证明了我们提出的方法的有效性。

关键词：联邦学习, 个性化, 协作

1 引言

现代学习任务通常由深度神经网络 (DNN) 实现，需要大量的训练数据才能达到满意的模型性能 [20] [18]。然而，由于数据量越来越大，收集数据的成本太高，甚至因为隐私保护而被禁止。因此，开发高效通信且保护隐私的学习算法对于充分利用客户端（例如数据孤岛和移动设备）中的数据具有重要意义 [25]。为此，联邦学习 (FL) 作为一种创新技术应运而生，无需收集原始数据即可对分散的客户端进行协作模型训练 [29]。典型的 FL 设置采用中央服务器来维护全局模型，并允许部分客户端参与不频繁的模型聚合，例如流行的 FedAvg，当客户端之间的本地数据独立且同分布 (IID) 时，它表现出了良好的性能。然而，在 FL 的背景下，客户端之间的数据分布通常不相同（非 IID 或异质性），因为不同的设备分别生成或收集数据，并且可能具有特定的偏好，包括特征分布漂移、标签分布倾斜和概念转移，这使得学习适用于所有客户的单一全局模型变得困难 [41] [21]。

为了解决这个问题，开发了个性化联合学习 (PFL)，其目标是为每个客户学习一个定制模型，该模型在本地数据上具有更好的性能，同时仍然受益于协作训练 [19] [36]。这种设置可以通过跨筒仓 FL 来激发，其中自主客户（例如医院和公司）可能希望满足客户特定的目标任务。实用的 FL 框架应该意识到数据的异构性，并在联合训练期间灵活地适应局部目标。另一方面，基于 DNN 的模型通常由用于从数据中提取低维特征嵌入的特征提取器和用于做出分类决策的分类器组成。深度学习在集中式系统和多任务学习中的成功表明，特征提取器起着通用结构的作用，而分类器往往与任务高度相关 [5] [8]。此外，实际 FL 问题中的客户端通常

处理类似的学习任务，并且在许多先前的工作中假设客户端之间的集群结构 [14] [34]。因此，学习更好的全局特征表示并利用局部任务之间的相关性对于改进个性化模型具有重要意义。

在这项工作中，主要考虑了标签分布转移场景，其中客户端之间的类数量相同，而每个类中的数据样本数量有明显的漂移，即本地任务的异构标签分布。作者通过利用共享表示和客户端间分类器协作，从多任务学习的角度研究联邦学习。具体来说，作者利用每个类的全局特征质心来规范局部训练，这可以被视为显式特征对齐，并且能够减少局部训练的特征提取器的表示多样性，从而促进全局聚合；还通过针对特定客户的线性组合进行灵活的分类器协作，这鼓励相似的客户更多地协作，并避免来自不相关客户的负迁移。为了估计适当的组合权重，利用局部特征统计数据 and 数据分布信息，通过解决二次规划问题来实现最佳偏差-方差权衡，从而最大限度地减少每个客户的预期测试损失。此外，通过稍加修改，所提出的框架仍然可以在概念转变场景下正常工作，在这种场景下，相同的标签可能在不同的客户中具有不同的含义。

2 相关工作

2.1 非独立同分布数据下的联邦学习

人们做出了许多努力来改进非独立同分布数据的 FL 全局模型学习。许多工作侧重于通过利用精心设计的目标正则化 [32] [1] [22] 和局部偏差校正 [17] 来优化局部学习算法。例如，FedProx [33] 在本地训练目标中添加了一个近端项，以保持更新的参数接近原始下载的模型，SCAFFOLD [17] 引入了控制变量来纠正本地训练目标中的漂移更新，MOON [23] 采用对比损失来改进表示学习。当客户端本地数据不平衡时，类平衡数据重采样和损失重加权方法可以提高训练性能 [6]。此外，还研究了数据共享机制和数据增强方法，以缓解非独立同分布数据挑战 [42] [13]。从模型聚合的角度来看，选择对全局模型性能贡献较大的客户端也可以加快收敛速度并减轻非独立同分布数据的影响 [37] [38]。随着公共数据的可用性，尽管存在数据异构性，但仍可以采用知识蒸馏技术来获得全局模型 [26] [43]。基于原型的方法也被用在一些 FL 工作中，例如 [30] 在全局聚合过程中提出了基于原型的权重注意力机制，以及 [31] [39] 利用原型来增强本地模型训练。与上述方法不同，本文旨在为每个客户学习定制模型。

2.2 个性化联邦学习

在文献中，流行的个性化 FL 方法包括执行局部和全局模式线性组合的加性模型混合，例如 L2CD [16] 和 APFL [11]；以及具有模型相异性惩罚的多任务学习，包括 FedMTL [35]、pFedMe [12] 和 Ditto [24]；利用特征提取器和分类器的参数解耦，例如 FedPer、LG-FedAvg 和 FedRep [3] [9]。一种特殊类型的个性化 FL 方法是集群 FL，它将相似的客户分组在一起并学习多个组内全局模型 [15] [27]。还针对细粒度联合研究了特定于客户端的模型聚合，例如 FedFomo 和 FedAMP [40] [4]，它们与作者的方法具有相似的想法。然而，现有的特定于客户的 FL 方法通常是通过以启发式方式评估模型相似性或验证准确性来开发的，这些技术需要在通信/计算开销和个性化有效性之间取得良好的平衡。基于高斯过程的 FedGP [2] 和基于选择性知识转移的解决方案也被开发出来，但是这些方法不可避免地依赖于公共共享数据集或诱导点集。此外，还研究了学习多个全局模型混合的 FedEM [28]，以便为每个客户端生成定

制模型。然而，pFedHN 要求每个客户端多次通信才能学习代表性嵌入，而 FedEM 显著增加了通信和计算/存储开销。最近，Fed-RoD [7] 提出使用平衡 softmax 来学习通用模型，使用 vanilla softmax 来学习个性化头部。作者的工作与 FedRep [10] 共享最相似的学习过程，但不同之处在于作者利用全局知识来指导局部表示学习，并为每个客户执行理论上保证的分类器头组合。

3 本文方法

3.1 本文方法概述

考虑采用 m 个客户端和一个中央服务器的设置，其中所有客户端都与服务器通信以协作训练个性化模型，而无需共享原始私有数据。在个性化 FL 中，每个客户端 i 在 $x \times y$ 上配备有自己的数据分布 $P_{XY}^{(i)}$ ，其中 x 是输入空间， y 是总共 K 个类别的标签空间。假设 $P_{XY}^{(i)}$ 和 $P_{XY}^{(j)}$ 对于任何一对客户端 i 和 j 都是不同的，这在 FL 中通常是这种情况。令 $\ell: x \times y \rightarrow R$ 表示给定局部模型 ω_i 和从 $P_{XY}^{(i)}$ 采样的数据点的损失函数，例如交叉熵损失，则 PFL 的底层优化目标可以形式化如下：

$$\min\{F(W) = \frac{1}{m} \sum_{i=1}^m E_{(x,y) \sim P_{XY}^{(i)}} [\ell(\omega_i; x, y)]\}$$

其中 $W = (\omega_1, \omega_2, \dots, \omega_m)$ 表示所有局部模型的集合。然而，真正的潜在分布是无法获得的，并且目标通常是通过经验风险最小化 (ERM) 来实现的。假设每个客户端可以访问从 $P_{XY}^{(i)}$ 采样的 n_i 个 IID 数据点，表示为 $D_i = \{(x_l^{(i)}, y_l^{(i)})\}_{l=1}^{n_i}$ ，其对应的经验分布为 $\hat{P}_Y^{(i)}$ ，并假设经验边际分布 $\hat{P}_Y^{(i)}$ 与真实的 $P_{XY}^{(i)}$ 相同。则训练目标为：

$$\omega^* = \operatorname{argmin}_w \frac{1}{m} \sum_{i=1}^m [\mathcal{L}_i(\omega_i; \Omega)]$$

其中 $\mathcal{L}_i(\omega_i) = \frac{1}{n_i} \sum_{l=1}^{n_i} \ell(\omega_i; x_j^{(i)}, y_j^{(i)})$ 是个人训练数据的局部平均损失，例如经验风险； Ω 是向相关客户端引入的某种全局信息， $\mathcal{R}_i(\cdot)$ 是预定义的正则化项，用于防止 ω_i 过度拟合。

3.2 共享特征表示

不失一般性，将深度神经网络解耦为表示层和最终决策层，其中前者也称为特征提取器，后者指分类任务中的分类器头。特征嵌入函数 $f: \mathcal{X} \rightarrow \mathbb{R}^d$ 是一个由 θ_f 参数化的可学习网络， d 是特征嵌入的维度。给定数据点 x ，提取的特征向量 $z = f(x)$ 的预测可以通过由 ϕ_g 参数化的线性函数 $g(z)$ 生成。在本文的其余部分中，为了简单起见，省略了 θ_f 和 ϕ_g 的下标，即 $\omega = \{\theta, \phi\}$ 。由于每个客户端的数据不足，本地学习的特征表示很容易过度拟合，从而不能很好地泛化。一个合理可行的想法是通过共享相同的特征表示层来利用其他客户端可用的数据。然而，对私有数据的多次本地更新将导致局部过拟合和跨客户端的高参数多样性，这使得聚合模型偏离最佳表示。为了解决这个问题，作者提出了一个用于局部特征表示学习的新正则化术语。

客户需要通过考虑监督学习损失和泛化误差来更新本地模型。为此，利用全局特征质心，并向局部训练目标引入新的正则化项，以使局部表示学习从全局数据中受益。局部正则化项

由下式给出:

$$\mathcal{R}_i(\theta_i; c) = \frac{\lambda}{n_i} \sum_{l=1}^{n_i} \frac{1}{d} \|f_{\theta_i}(x_l) - c_{yl}\|^2$$

其中 $f_{\theta_i}(x_l)$ 是给定数据点 x_j 的局部特征嵌入, c_{yj} 是类 y_j 对应的全局特征质心, λ 是平衡监督损失和正则化损失的超参数。通过利用全局语义特征信息, 这样的正则化项对每个客户都有显着的好处。直观上, 它使每个客户端能够通过显式的特征分布对齐来学习任务不变的表示。因此, 局部特征提取器 $\{\theta_i\}_{i=1}^m$ 的多样性也可以被正则化, 同时最小化局部分类误差。能够注意到, 全局特征质心类似于原型的概念, 广泛应用于少样本学习、对比学习和领域适应; 该正则化项可以通过显式减少泛化误差来启发, 从而有助于提高测试准确性。

3.3 分类器协作

除了通过共享表示层来改进特征提取器之外, 作者认为合并来自具有相似数据分布的其他客户端的分类器也可以提供性能增益。与之前仅维护本地训练的分类器的工作不同, 作者还对分类器进行特定于客户端的加权平均值以获得更好的本地分类器。直观上, 当本地数据不足时, 本地学习的分类器可能具有较高的方差, 因此共享相似数据分布的那些客户端实际上可以通过客户端间的知识迁移来协作训练个性化分类器。挑战在于如何评估客户之间的相似性和可转移性。为此, 对每个客户端 i 接收到的分类器进行线性组合, 以减少本地测试损失:

$$\hat{\phi}_i^{(t+1)} = \sum_{j=1}^m \alpha_{ij} \phi_j^{(t+1)}, s.t. \sum_{j=1}^m \alpha_{ij} = 1$$

通过最小化局部预期测试损失来确定每个系数 $\alpha_{ij} \geq 0$, 可以将其表示为以下优化问题:

$$\alpha_i^* = \underset{\alpha_i}{\operatorname{argmin}} \mathbb{E}_{(x,y) \sim P_{XY}^{(i)}} [\ell(\theta, \sum_{j=1}^m \alpha_{ij} \phi_j x, y)]$$

为了更好的协作, 需要在训练过程中自适应地更新系数 α_i 。

考虑一个简单的情况, 其中数据在客户端之间是同质的。在这种情况下, 学习单个全局模型对于泛化来说是最佳的。然而, 一些现有的方法, 例如 FedPer 和 FedRep, 没有考虑分类器头的组合, 因此在异构性较低的情况下导致性能有限。相比之下, 作者的方法进行自适应分类器组合, 并且在同质和异构场景中都有效。

4 算法设计

在本节中, 作者设计了一种交替优化方法来迭代学习局部分类器和全局特征提取器。本地更新的模型参数以及特征统计数据将传输到中央服务器进行聚合。

4.1 本地训练过程

对于第 t 轮的局部模型训练, 我们首先用接收到的全局聚合 $\theta_i^{(t)}$ 替换局部表示层 $\tilde{\theta}^{(t)}$, 并类似地更新私有分类器。然后, 执行随机梯度下降步骤来训练模型参数的两部分, 如下所示:

步骤 1: 固定 θ_i , 更新 ϕ_i 。在本地数据上在一个 epoch 内使用梯度下降训练 ϕ_i :

$$\phi_i^{(t)} \leftarrow \phi_i^{(t)} - \eta_g \nabla_{\phi} \ell(\theta_i^{(t)}, \phi_i^{(t)}; \xi_i)$$

其中 ξ_i 代表数据的小批量, η_g 是用于更新分类器的学习率

步骤 2: 固定新的 ϕ_i , 更新 ϕ_i 。获得新的局部分类器后, 基于多个 epoch 的私有数据和全局质心来训练局部特征提取器:

$$\theta_i^{(t)} \leftarrow \theta_i^{(t)} - \eta_g \nabla_{\theta} [\ell(\theta_i^{(t)}, \phi_i^{(t+1)}; \xi_i) + \mathcal{R}_i(\theta_i^{(t)}; c^{(t)})]$$

其中 η_f 是更新表示层的学习率, $c^{(t)} \in \mathbb{R}^{K \times d}$ 是每个类的全局特征质心向量的集合, $K = |\mathcal{I}|$ 是类的总数。

特征统计提取: 在局部特征提取器更新之前, 每个客户端应通过局部数据集的单遍提取局部特征统计 $\hat{V}_i^{(t)}$ 和 $V_i^{(t)}$, 这将用于估计最佳分类器组合权重每个客户。此外, 更新局部特征提取器后, 按以下公式计算每个类的局部特征质心:

$$\hat{c}_{i,k}^{(t+1)} = \frac{\sum_{l=1}^{n_i} 1(y_l^{(i)} = k) f_{\theta_i^{(t+1)}}(x_l^{(i)})}{\sum_{l=1}^{n_i} 1(y_l^{(i)} = k)}, \forall k \in [K]$$

4.2 全局聚合

全局特征表示: 与常见算法一样, 服务器对本地表示层进行加权平均, 每个系数由本地数据大小确定:

$$\tilde{\theta}^{(t+1)} = \sum_{i=1}^m \beta_i \theta_i^{(t)}, \beta_i = \frac{n_i}{\sum_{i=1}^m n_i}$$

分类器组合: 服务器使用接收到的特征统计数据更新组合权重向量 α_i , 并对每个客户端 i 进行分类器组合。

更新全局特征质心: 接收到局部特征质心后, 进行以下质心聚合操作以生成每个类 k 的估计全局质心 c_k :

$$c_k^{(t+1)} = \frac{1}{\sum_{i=1}^m n_{i,k}} \sum_{i=1}^m n_{i,k} \hat{c}_{i,k}^{(t+1)}, \forall k \in [K]$$

4.3 与已有开源代码对比

没有参考任何相关源代码

5 实验结果分析

5.1 实验设置

数据集: CIFAR-10

数据分割: 使所有客户端具有相同的数据大小, 其中 $s\%$ 的数据 (默认为 20%) 从所有类中统一采样, 其余 $(100 - s)\%$ 来自每个客户端的一组主导类。并且保持本地训练数据的大小较小, 以满足 FL 的需要。每个客户端的测试数据与训练数据具有相同的分布。

比较方法: 比较以下基线: Local-only, 每个客户端在本地训练其模型、FedAvg

训练设置: 采用小批量 SGD 作为所有方法的局部优化器, 除非明确指定, 否则局部训练时期的数量设置为 $E = 5$ 。所有数据集的全局通信轮数设置为 200, 其中所有 FL 方法通过更多通信几乎没有或没有精度增益。

5.2 实验结果

训练轮次比较：在两种设置上进行实验，其中本地 epoch 分别为 1 和 5。主要结果如图 1 所示，可以看出当客户端本地训练轮数设置为 5 时，模型准确率比设置为 1 时高。这可能是因为较小的本地训练轮数会导致每个客户端的本地更新不足，从而影响整体模型的性能



图 1. 训练轮次对模型准确率的影响

FedPAC 与 FedAvg 比较：令训练时所用的数据集、全局通信轮数与本地训练轮数等其他参数均保持一致。可以从图 2 看出 FedPAC 的测试准确率高于 FedAvg：

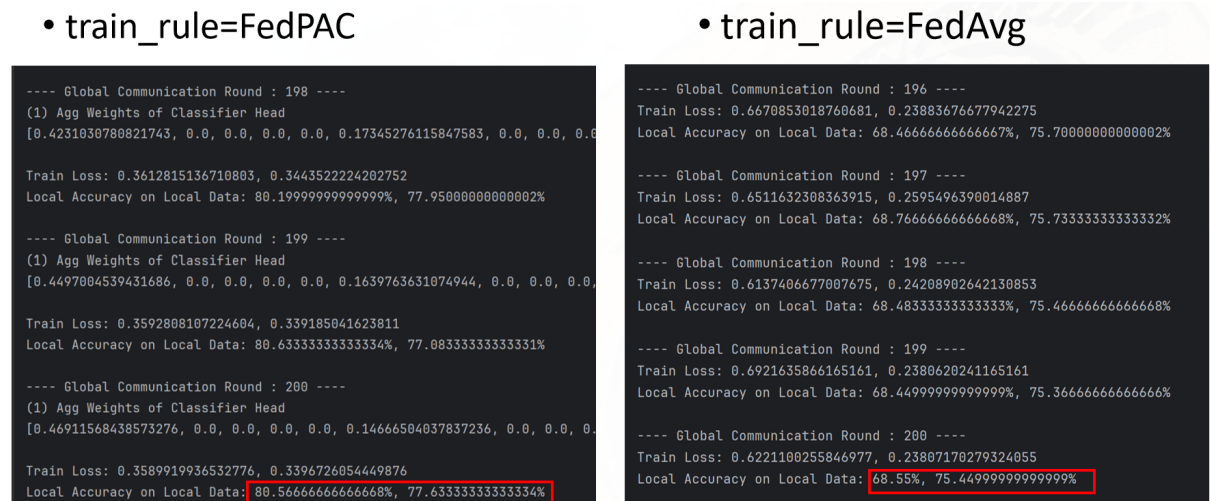


图 2. FedPAC 与 FedAvg 的测试准确率的比较

FedPAC 与本地训练比较：令训练时所用的数据集、全局通信轮数与本地训练轮数等其他参数均保持一致。可以从图 3 看出 FedPAC 的测试准确率高本地训练模型：

• train_rule=FedPAC

```

---- Global Communication Round : 198 ----
(1) Agg Weights of Classifier Head
[0.4231030780821743, 0.0, 0.0, 0.0, 0.0, 0.17345276115847583, 0.0, 0.0, 0.0]

Train Loss: 0.3612815136710803, 0.3443522224202752
Local Accuracy on Local Data: 80.19999999999999%, 77.95000000000002%

---- Global Communication Round : 199 ----
(1) Agg Weights of Classifier Head
[0.4497004539431686, 0.0, 0.0, 0.0, 0.0, 0.1639763631074944, 0.0, 0.0, 0.0]

Train Loss: 0.3592808107224604, 0.339185041623811
Local Accuracy on Local Data: 80.63333333333334%, 77.08333333333331%

---- Global Communication Round : 200 ----
(1) Agg Weights of Classifier Head
[0.46911568438573276, 0.0, 0.0, 0.0, 0.0, 0.14666504037837236, 0.0, 0.0, 0.0]

Train Loss: 0.3589919936532776, 0.3396726054449876
Local Accuracy on Local Data: 80.56666666666668%, 77.63333333333334%

```

• train_rule=Local

```

---- Global Communication Round : 196 ----
Train Loss: 0.006446111105469754, 0.0012587458317284473
Local Accuracy on Local Data: 65.63333333333334%, 65.66666666666669%

---- Global Communication Round : 197 ----
Train Loss: 0.001542480858915951, 0.0024191414326196537
Local Accuracy on Local Data: 65.66666666666669%, 65.21666666666667%

---- Global Communication Round : 198 ----
Train Loss: 0.014512810987071134, 0.0013521904504159466
Local Accuracy on Local Data: 65.21666666666667%, 65.35000000000001%

---- Global Communication Round : 199 ----
Train Loss: 0.0018633508778293617, 0.0031009098835056647
Local Accuracy on Local Data: 65.35000000000001%, 65.45%

---- Global Communication Round : 200 ----
Train Loss: 0.007046195695875212, 0.001758131275710184
Local Accuracy on Local Data: 65.45%, 65.55000000000001%

```

图 3. FedPAC 与本地模型的测试准确率的比较

FedPAC 在平均测试精度方面优于本地训练模型，这证明了全局特征对齐和客户端间分类器协作的有效性和优势。

6 总结与展望

在本文中，介绍了用于增强表示学习的全局特征对齐方法，以及在联邦学习（FL）中构建个性化分类器的新型分类器组合算法。这些方法在理论和实验上都证明了其在异构环境中的实用性。未来的工作将包括在更复杂的环境中分析最佳的模型个性化策略，例如在具有动态数据分布的去中心化系统或客户端场景中，并探索本地特征提取器的最优聚合方法。这将有助于进一步提升模型在不同环境下的适应性和效率，以及在保护用户隐私的同时提高模型的整体性能。

参考文献

- [1] Durmus Alp Emre Acar, Yue Zhao, Ramon Matas Navarro, Matthew Mattina, Paul N. Whatmough, and Venkatesh Saligrama. Federated learning based on dynamic regularization. *ArXiv*, abs/2111.04263, 2021.
- [2] Idan Achituve, Aviv Shamsian, Aviv Navon, Gal Chechik, and Ethan Fetaya. Personalized federated learning with gaussian processes. In *Neural Information Processing Systems*, 2021.
- [3] Manoj Ghuhana Arivazhagan, V. Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *ArXiv*, abs/1912.00818, 2019.
- [4] Martin Beaussart, Felix Grimberg, Mary-Anne Hartley, and Martin Jaggi. Waffle: Weighted averaging for personalized federated learning. *ArXiv*, abs/2110.06978, 2021.

- [5] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, 2013.
- [6] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning. *ArXiv*, abs/2107.00778, 2021.
- [7] Hong-You Chen and Wei-Lun Chao. On bridging generic and personalized federated learning for image classification. In *International Conference on Learning Representations*, 2022.
- [8] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. *ArXiv*, abs/2102.07078, 2021.
- [9] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. *ArXiv*, abs/2102.07078, 2021.
- [10] Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. *ArXiv*, abs/2102.07078, 2021.
- [11] Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive personalized federated learning. *ArXiv*, abs/2003.13461, 2020.
- [12] Canh T. Dinh, Nguyen Hoang Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. *ArXiv*, abs/2006.08848, 2020.
- [13] Yann Fraboni, Richard Vidal, Laetitia Kamani, and Marco Lorenzi. Clustered sampling: Low-variance and improved representativity for clients selection in federated learning. *ArXiv*, abs/2105.05883, 2021.
- [14] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. *IEEE Transactions on Information Theory*, 68(12):8076–8091, 2022.
- [15] Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. *IEEE Transactions on Information Theory*, 68(12):8076–8091, 2022.
- [16] Filip Hanzely and Peter Richtárik. Federated learning of a mixture of global and local models. *ArXiv*, abs/2002.05516, 2020.
- [17] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank J. Reddi, Sebastian U. Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, 2019.
- [18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, 60(6):84–90, may 2017.

- [19] Viraj Kulkarni, Milind Kulkarni, and Aniruddha Pant. Survey of personalization techniques for federated learning. In *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, pages 794–797, 2020.
- [20] Bengio Y. Hinton G LeCun, Y. Deep learning. *Nature*, 521(11):436–444, 2015.
- [21] Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, pages 965–978, 2022.
- [22] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10708–10717, 2021.
- [23] Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10708–10717, 2021.
- [24] Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *International Conference on Machine Learning*, 2020.
- [25] Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3):50–60, 2020.
- [26] Tao Lin, Lingjing Kong, Sebastian U. Stich, and Martin Jaggi. Ensemble distillation for robust model fusion in federated learning. *ArXiv*, abs/2006.07242, 2020.
- [27] Y. Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. Three approaches for personalization with applications to federated learning. *ArXiv*, abs/2002.10619, 2020.
- [28] Othmane Marfoq, Giovanni Neglia, Aurélien Bellet, Laetitia Kameni, and Richard Vidal. Federated multi-task learning under a mixture of distributions. In *Neural Information Processing Systems*, 2021.
- [29] H. B. McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *International Conference on Artificial Intelligence and Statistics*, 2016.
- [30] Umberto Michieli and Mete Ozay. Prototype guided federated learning of visual feature representations. *ArXiv*, abs/2105.08982, 2021.
- [31] Xutong Mu, Yulong Shen, Ke Cheng, Xueli Geng, Jiaxuan Fu, Tao Zhang, and Zhiwei Zhang. Fedproc: Prototypical contrastive federated learning on non-iid data. *Future Generation Computer Systems*, 143:93–104, 2023.

- [32] Anit Kumar Sahu, Tian Li, Maziar Sanjabi, Manzil Zaheer, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *arXiv: Learning*, 2018.
- [33] Anit Kumar Sahu, Tian Li, Maziar Sanjabi, Manzil Zaheer, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. *arXiv: Learning*, 2018.
- [34] Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8):3710–3722, 2021.
- [35] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet Talwalkar. Federated multi-task learning. In *Neural Information Processing Systems*, 2017.
- [36] Alysa Ziyang Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards personalized federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, page 1–17, 2022.
- [37] Minxue Tang, Xuefei Ning, Yitu Wang, Jingwei Sun, Yu Wang, Hai Li, and Yiran Chen. Fedcor: Correlation-based active client selection strategy for heterogeneous federated learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10092–10101, 2022.
- [38] Hongda Wu and Ping Wang. Node selection toward faster convergence for federated learning on non-iid data. *IEEE Transactions on Network Science and Engineering*, 9(5):3099–3111, 2022.
- [39] Rui Ye, Zhenyang Ni, Chenxin Xu, Jianyu Wang, Siheng Chen, and Yonina C. Eldar. Fedfm: Anchor-based feature matching for data heterogeneity in federated learning. *IEEE Transactions on Signal Processing*, 71:4224–4239, 2023.
- [40] Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and José Manuel Álvarez. Personalized federated learning with first order model optimization, 2020.
- [41] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *ArXiv*, abs/1806.00582, 2018.
- [42] Yue Zhao, Meng Li, Liangzhen Lai, Naveen Suda, Damon Civin, and Vikas Chandra. Federated learning with non-iid data. *ArXiv*, abs/1806.00582, 2018.
- [43] Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-free knowledge distillation for heterogeneous federated learning. *Proceedings of machine learning research*, 139:12878–12889, 2021.